



HAL
open science

Caractérisation in silico des bactéries des cultures de la microalgue marine *Haslea ostrearia*

Aurélie Peticca

► To cite this version:

Aurélie Peticca. Caractérisation in silico des bactéries des cultures de la microalgue marine *Haslea ostrearia*. Génétique des plantes. Le Mans Université, 2023. Français. NNT : 2023LEMA1023 . tel-04465852

HAL Id: tel-04465852

<https://theses.hal.science/tel-04465852v1>

Submitted on 19 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT

DE
LE MANS UNIVERSITE
SOUS LE SCEAU DE
LA COMUE ANGERS – LE MANS

ECOLE DOCTORALE N° 642
Ecole doctorale Végétal, Animal, Aliment, Mer, Environnement
Spécialité : Génétique, Génomique et Bio-informatique

Par

Aurélie PETICCA

« **Caractérisation *in silico* des bactéries des cultures de la microalgue marine *Haslea ostrearia* »**

Thèse présentée et soutenue au Mans, le 11/12/2023

Unité de recherche : **Biologie des Organismes, Stress, Santé, Environnement (BIOSSE)**

Thèse N° : **2023LEMA1023**

Rapporteurs avant soutenance :

François-Yves BOUGET
Jonathan FILEE

Directeur de Recherche au CNRS (Banyuls-sur-mer)
Chargé de Recherche au CNRS (Gif sur Yvette)

Composition du Jury :

Président : François-Yves BOUGET
Examineurs : Anne-Laure BANULS
Agnès PETIT
Jonathan FILEE
Directrice de thèse : Nathalie CASSE
Encadrants de thèse : Benoît CHENAIS
François SABOT

Directeur de Recherche au CNRS (Banyuls-sur-mer)
Directrice de Recherche à l'IRD (Montpellier)
Maître de Conférence au CEPR (Tours)
Chargé de Recherche au CNRS (Gif sur Yvette)
Maître de Conférence au BiOSSE (Le Mans)
Professeur au BiOSSE (Le Mans)
Directeur de Recherche à l'IRD (Montpellier)

Remerciements

Il me sera difficile de remercier toutes les personnes qui ont joué un rôle au cours de ma thèse en raison de leur nombre important, mais je vais tout de même tenter de le faire !

Pour commencer, je souhaiterais remercier ceux sans qui cette thèse n'aurait jamais été possible.

Dr Nathalie Casse, ma directrice de thèse, merci de t'être assurée du bon déroulement de la thèse afin que je ne manque de rien durant celle-ci. Pr Benoît Chénaïs, mon premier co-encadrant, merci pour tes conseils quant aux implications que représentent une thèse et sur la vie "après soutenance". Dr François Sabot, mon deuxième co-encadrant, merci pour nos nombreuses réunions et discussions scientifiques (ou non) mais également pour ton honnêteté en toutes circonstances. J'espère sincèrement que notre collaboration pourra perdurer dans le futur, ou pouvoir te croiser dans un cadre professionnel ou plus informel.

A vous qui avez su m'accompagner et me guider tout le long de ce parcours de thèse, un immense merci.

Je tenais également à remercier la région Pays de la Loire et Le Mans métropole pour avoir financé le pari scientifique LANCOM dont est issu mon sujet de thèse.

Je remercie le laboratoire BioSSE pour cette expérience qui m'imprènera pendant de nombreuses années, et dictera mes choix futurs. J'y joins un merci particulier au Dr Héléne Gâteau ainsi qu'au Dr Mostefa Fodil pour avoir pris soin des cultures de *Haslea ostrearia* et avoir extrait son ADN (et celui de ses copines bactéries). Mais surtout merci de m'avoir permis de rencontrer d'admirables personnes, et fervents défenseurs de brookies : Dr Julie Séveno, Dr Mariem Zaidi, le (grand) technicien Romuald Leroux et Dr Maxime Louzon. Une super équipe pour un super bureau, bien que quelqu'un (dont on taira le nom) a osé se l'approprier comme salle de pause. Sans votre aide les journées auraient été bien plus longues, c'est Scientifiqu'Mans prouvé !

Également un immense merci aux personnes extérieures au laboratoire que j'ai pu côtoyer au cours de ma thèse. Christophe Klopp (Genotoul), Dr Juliette Hayer (IRD), Dr Cécile Triay (IRD), Julie Orjuela-Bouniol (IRD), Aurore Comte (IRD), ainsi que l'ensemble des membres de l'IRD de Montpellier. Je ne pourrai jamais vous remercier assez pour votre bonne humeur et votre vision de la recherche. Ce fut un véritable honneur d'être accueillie avec autant de bienveillance dans les locaux de l'IRD de Montpellier, parmi l'équipe DIADE.

Je vous suis reconnaissante de m'avoir offert l'opportunité d'échanger et de travailler avec vous tous.

Mais que serait une thèse sans le soutien des amis et de la famille ? Merci pour cet encouragement incommensurable même si pour vous le sujet de recherche a quelques fois été incompréhensible, voire d'un autre monde.

Merci à mes parents et mon frère d'avoir si souvent entendu mes plaintes, mais également d'avoir su fêter chacune de mes petites victoires. Merci également à ma grand-mère, ma tante et toute ma famille qui, bien que je n'ai pas pu les voir autant que désiré, ont toujours été à mes côtés. Si les collègues de bureau ont su accélérer mes journées, c'est vous qui m'avez permis de les vivre et d'arriver jusqu'ici.

J'aimerais également remercier la très bientôt Dr Anna Tran qui, au moment où j'écris ces mots, devrait passer sa soutenance prochainement. Merci pour tous ces échanges durant nos thèses respectives et le soutien moral qui en a découlé. Souris ou microalgue, même combat ! J'ai grandement adoré ma période à Montpellier grâce à toi.

For the last but not the least, j'aimerais remercier mon compagnon de tous les jours, Théophile. Ton courage pour braver la "fiabilité sans faille" du système ferroviaire français, et ton oreille attentive m'ont souvent permis de poursuivre malgré les obstacles. A toi, qui termine ce voyage à mes côtés, merci.

Merci à tous et pour tout,
- Aurélie Peticca

Résumé

Haslea ostrearia est une algue microscopique marine provenant de la famille des diatomées, particulièrement connue pour le pigment qui lui est associé, la marennine. Cette microalgue est fréquemment retrouvée dans les bassins d'huîtres de la côte Atlantique française, qu'elle teinte d'une couleur verdâtre. Bien que présentant de forts intérêts économiques, les huîtres "vertes" étant plus chères, et médicaux, pour les activités antioxydantes de la marennine, *H. ostrearia* reste encore très méconnue. Son écologie complexe se traduit entre autres par l'incapacité d'obtenir des cultures "pures" (axéniques), ce qui contamine tout séquençage d'ADN par celui de bactéries associées, et empêche ainsi la reconstitution de son génome. Dans le cadre de cette thèse, nous nous proposons d'étudier les données ADN obtenues des cultures d'une souche de *H. ostrearia* afin d'identifier et étudier les bactéries associées à celle-ci.

Dans un premier temps, les données de séquençage ADN des cultures ont été assemblées et triées pour identifier les séquences appartenant à la microalgue, ce qui a permis d'assembler ses génomes chloroplastiques et mitochondriaux. Devant la faible quantité de ces dernières, il a alors été proposé une identification précise des bactéries présentes dans les cultures en étudiant la région 16S V3V4, utilisée régulièrement par d'autres analyses de ce type. Dans un dernier temps, l'étude des génomes bactériens entier a permis de proposer des pistes pour la création d'un nouveau protocole de culture afin de réduire la diversité bactérienne à son minimum. L'utilisation actuelle d'antibiotiques à large spectre ne semble en effet pas être suffisante, notamment suite à la présence de nombreux gènes d'antibiorésistance chez les bactéries.

L'ensemble de ces travaux représente une étape importante pour l'obtention du génome de *H. ostrearia*, par l'acquisition de cultures "propres" et reproductibles (cultures avec une population bactérienne connue, contrôlée et réduite, voire inexistante). Ils ouvrent ainsi la voie à de nombreuses études sur les interactions entre ces bactéries et *H. ostrearia*, mais également sur la présence ou non de ces dernières dans chaque milieu naturel de la microalgue, et l'évolution de leur concentration au fil de ses cycles ou de son temps de culture.

Sommaire

Remerciements	2
Résumé	4
Sommaire	5
Liste des Productions Scientifiques	8
Liste des Figures	9
Liste des Tableaux	10
Résumé étendu	11
1. Organisme d'intérêt.....	11
2. Travaux réalisés durant la thèse	12
3. Conclusion et perspectives	14
Introduction	16
Partie 1 : la diatomée <i>Haslea ostrearia</i>	16
1. Morphologie des diatomées	17
2. Origines évolutives des diatomées.....	19
3. Classification taxonomique des diatomées	23
4. Cycle de vie des diatomées	26
5. Interactions biotiques des diatomées	29
6. Les caractéristiques du sous-genre <i>Haslea</i> bleue.....	33
7. La génomique de <i>Haslea ostrearia</i>	37
Partie 2 : L'assemblage de génomes	41
1. Histoire du séquençage de l'ADN	41
2. Les techniques de séquençage	43
3. Les méthodes d'assemblage et de métassemblage.....	47
4. L'élimination des génomes "contaminants", une des limites de la métagénomique	52
Partie 3 : L'antibiorésistance, un problème mondial	56
1. Depuis quand parle-t-on d'antibiorésistance ?	56
2. Les enjeux de l'antibiorésistance	57
3. La recherche des marqueurs d'une potentielle résistance aux antibiotiques	60

Chapitre 1 : A la recherche du génome de <i>H. ostrearia</i>	63
1. Contexte des travaux de recherche	63
2. Aperçu général des données ONT et Illumina ainsi que du métagénome obtenu	65
3. Article sur les génomes chloroplastique et mitochondrial de <i>H. ostrearia</i>	71
4. Conclusion du chapitre et prélude du chapitre suivant	85
Chapitre 2 : Identifier les bactéries présentes dans les cultures	86
1. Contexte des travaux de recherche	86
2. Article sur l'identification par 16S des bactéries présentes dans les cultures de <i>H. ostrearia</i>	87
3. Conclusion du chapitre	105
Chapitre 3 : Etude des gènes de résistance aux antibiotiques chez les bactéries des cultures de <i>H. ostrearia</i>	107
1. Contexte des travaux de recherche	107
2. Article sur la recherche de AMRs et ARGs	108
3. Résultats complémentaires.....	134
4. Conclusion du chapitre	138
Conclusion générale	140
1. Remise en contexte des résultats	140
2. La place de la métagénomique dans l'amélioration et la compréhension des manipulations biologiques.....	142
3. Perspectives proches et lointaines.....	143
Bibliographie	146
Annexes	169
1. Tentatives approfondies de décontamination du métagénome	169
1) Analyses des contigs > 1Mb.....	169
2) Recherche des contigs appartenant à <i>H. ostrearia</i> dans le métagénome.....	175
2. Supplementary data of "Complete mitochondrial genome and draft chloroplastic genome of <i>Haslea ostrearia</i> (Simonsen 1974)"	181

3. Supplementary data of “To improve the axenic culture protocol, the bacteria present in <i>Haslea ostrearia</i> cultures need to be studied”	182
References	188
4. Supplementary data of “Identification of antibiotic-resistance genes in order to reduce the bacterial content of <i>Haslea ostrearia</i> cultures”	191

Liste des Productions Scientifiques

- **Articles scientifiques**

Mitochondrial DNA Part B: Resources [Research Paper]: **Peticca A**, Fodil M, Gateau H, Mouget JL, Sabot F, Chénaïs B, Casse N (2023) *Complete mitochondrial genome and draft chloroplastic genome of Haslea ostrearia (Naviculaceae, Bacillariophyceae)*. Disponible sur: <https://doi.org/10.1080/23802359.2023.2268747>

Marine Drugs [Research Paper]: Gabed N, Verret F, **Peticca A**, Kryvoruchko I, Gastineau R, Bosson O, Séveno J, Davidovich O, Davidovich N, Witkowski A, Kristoffersen JB, Benali A, Ioannou E, Koutsaviti A, Roussis V, Gâteau H, Phimmaha S, Leignel V, Badawi M, Khiar F, Francezon N, Fodil M, Pasetto P, Mouget J-L (2022) *What Was Old Is New Again: The Pennate Diatom Haslea ostrearia (Gaillon) Simonsen in the Multi-Omic Age*. *Mar. Drugs*, 20, 234. Disponible sur : <https://doi.org/10.3390/md20040234>

Scientific Data [Research Paper]: **Peticca A**, Fodil M, Chénaïs B, Sabot F, Casse N (2023) *The need to stabilize the bacterial community present in the cultures of the marine diatom Haslea ostrearia*. En relecture par les paires, disponible sur: <https://doi.org/10.21203/rs.3.rs-2751923/v1>

PCI Microbiol [Research Paper]: **Peticca A**, Chénaïs B, Casse N, Sabot F (2023) *Antibiotics potentially capable of reducing the bacterial content of Haslea ostrearia cultures*. Manuscrit prochainement soumis.

- **Communications orales (Poster)**

Les Journées Ouvertes en Biologie, Informatique et Mathématiques JOBIM 2023 (Nice), 27/06/2023 : **Peticca A**, Klopp C, Fodil M, Chénaïs B, Casse N, Sabot F (2023) *Identification of a microbiome, the advantages of the metagenomic method over the classical 16S method*

Concours Ma thèse en 180 secondes (Le Mans), 10/03/2023 : **Peticca A** (2023)

Concours Exposé ta thèse (Le Mans), 18/11/2022 : **Peticca A** (2022) *Haslea, à l'assemblage !*

Liste des Figures

Figure 0-1 : Exemples des structures de valves retrouvées chez différentes espèces de diatomées à symétrie radiale et pennée	18
Figure 0-2 : Représentation schématique de la paroi cellulaire des diatomées.....	18
Figure 0-3 : Schéma des deux hypothèses alternatives sur l'origine du chloroplaste des diatomées et de sa structure	22
Figure 0-4 : Représentation des différents critères utilisés pour la classification des diatomées	26
Figure 0-5 : Schématisation des phases de reproduction végétatives (A) et sexuées (B) des diatomées.....	28
Figure 0-6 : Schématisation des types d'interactions possibles entre espèces	33
Figure 0-7 : Vue microscopie de <i>Haslea ostrearia</i> (A), vue d'une huître verdie par la présence de marennine sur ses branchies (B) et vue d'un bloom de <i>Haslea ostrearia</i> (C).....	36
Figure 0-8 : Localisation des <i>Haslea ostrearia</i> à travers le monde d'après la littérature.....	37
Figure 0-9 : Courbe de l'évolution du prix d'un séquençage du génome humain de 2001 à 2022	43
Figure 0-10 : Représentation schématique des méthodes de séquençage NGS (A) et TGS (B)	46
Figure 0-11 : Représentation schématique des k-mers extraits d'une séquence ADN (A) et du graphe de Bruijn correspondant (B).....	49
Figure 0-12 : Exemple d'un graphe de répétition d'un assemblage bactérien réalisé par le logiciel Flye	51
Figure 0-13 : Représentation schématique des ARNs ribosomiques 16S (A) et 18S (B)	55
Figure 0-14 : Frise chronologique des évènements marquants dans l'histoire du phénomène d'antibiorésistance depuis la création des antibiotiques.....	57

Figure 0-15 : Schéma explicatif du fonctionnement du logiciel RGI et des résultats obtenus avec l'utilisation de la base de données CARD	62
Figure 1-1 : Courbe modélisant le contenu en Guanine et Cytosine moyen (%GC) des lectures Illumina séquencées des cultures de <i>Haslea ostrearia</i>	67
Figure 1-2 : Estimation de l'hétérozygotie, du contenu répété et de la taille des génomes contenus dans les lectures de séquençage Illumina des cultures de <i>Haslea ostrearia</i>	68
Figure 1-3 : Résumé des statistiques obtenus sur le métagénome obtenu à partir des séquençages ADN des cultures de <i>Haslea ostrearia</i>	70
Figure 2-1 : Représentation schématique des régions variables et hypervariables de l'ARN ribosomique 16S et des régions séquencées pour l'étude	104

Liste des Tableaux

Tableau 0-1 : Tableau récapitulatif des caractéristiques des génomes de diatomées actuellement disponibles	39
Tableau 0-2 : Tableau récapitulatif des informations sur les principales techniques de séquençage utilisées en Septembre 2023	47
Tableau 1-1 : Séquençages réalisés sur l'isolat NCC 532 de <i>Haslea ostrearia</i> au laboratoire BiOSSE entre 2020 et 2021 inclus	63
Tableau 1-2 : Liste des amorces utilisées pour la recherche in silico des gènes ARNr 16S dans le métagénome de <i>Haslea ostrearia</i>	172
Tableau 1-3 : Informations sur les résultats obtenus pour l'analyse taxonomique des contigs dépassant la taille de 1 Mb et appartenant à l'assemblage métagénomique des cultures de <i>Haslea ostrearia</i>	174
Tableau 1-4 : Informations sur les séquences identifiées comme susceptibles d'appartenir au génome de <i>Haslea ostrearia</i>	180
Tableau 3-1 : Informations sur les plasmides identifiées dans le métagénome obtenu pour les cultures de <i>Haslea ostrearia</i> et les gènes de résistance aux antibiotiques ou aux antimicrobiens retrouvés à l'intérieur de leur séquence	137

Résumé étendu

1. Organisme d'intérêt

Les diatomées sont des algues microscopiques très largement étudiées suite à leur rôle écologique important. En effet, ces organismes eucaryotes unicellulaires évoluent dans les milieux aquatiques et sont capables de photosynthèse, responsables d'environ 40% de la production primaire d'oxygène des océans elles ne représentent pourtant que 1% de la biomasse photosynthétique terrestre (Marshall et Orr 1928, Moore et al 2001, Fu et al 2022). Les diatomées sont retrouvées dans des zones tropicales à polaires, dans des eaux douces ou salines (Vilmi et al 2015, Malviya et 2016, Otte et al 2023), ainsi qu'au fond des eaux (benthiques) ou en suspension (planctoniques) (Hu et al 2022). Ayant pu s'adapter à autant d'environnements différents, il n'est donc pas surprenant d'apprendre que les diatomées, également appelées *Bacillariophyta*, jouissent d'une très grande diversité en termes d'espèces (Malviya et al 2016).

En plus de leur activité photosynthétique, il a été découvert que ces microalgues impactaient grandement les réseaux trophiques de leurs environnements en influant sur plusieurs cycles géochimiques: Azote, Carbone, Fer, Silice, Phosphate, etc (Allen et al 2006, Fry et Wainright 1991, Muggli et al. 1996, Ragueneau et al 2006, Sun et al 2011). C'est pourquoi il a été fait état de nombreuses utilisations par l'être humain de ces microalgues au fil des années, et ce dans de nombreux domaines tels que l'industrie (Zahajská 2020), l'agroalimentaire (Lebeau et Robert 2003) ou le pharmaceutique (Nieri et al 2023).

Les travaux de thèse présentés ici s'articulent autour de la diatomée de la famille des *Naviculaceae*, nommée *Haslea (H.) ostrearia* (Simonsen 1974). La grande particularité de cette espèce est la capacité de synthétiser un pigment supplémentaire de ceux spécifiques au phylum (Mouget et al 1999) : la marennine (Prasetyia et al 2020). Bien que connu depuis 1886 (Moreau 1967), ce pigment bleuté reste cependant très incompris à ce jour et ce

indépendamment des nombreuses études sur le sujet. Par exemple, la nature chimique de la marennine reste inconnue et seules quelques suppositions sur sa complexité ont été publiées (Zebiri et al 2023, Pouvreau et al 2006). Son origine est également indéterminée puisqu'aucune description des mécanismes impliqués dans sa synthèse n'existe actuellement. Cependant, ce pigment possède un intérêt en agroalimentaire et en cosmétique, pouvant à terme remplacer des colorants pétrochimiques (Sigurdson et al 2017), mais également en aquaculture, grâce à ses bienfaits sur la croissance des larves de bivalves (Latour et al 2023).

A l'instar de la marennine, peu de connaissances sont disponibles sur *H. ostrearia*, la majorité des études réalisées se concentrant sur son pigment (Pouvreau 2006, Gastineau et al 2012, Falaise et al 2019, Méresse et al 2023). Néanmoins, des travaux publiés en 2016 font référence à un potentiel microbiome associé à une amélioration de la croissance de la microalgue lors de sa culture (Lépinay et al 2016, 2018). Ces bactéries pourraient être la raison derrière l'absence d'une reconstitution simple du génome de *H. ostrearia*, et de la seule publication des génomes de ses organelles (Dąbek et al 2022).

2. Travaux réalisés durant la thèse

L'objectif initial était de reconstituer le génome de *H. ostrearia* afin d'en étudier la composition en longs ARNs non codants. Cependant, de nombreux ADN bactériens ont été retrouvés dans les séquençages des cultures de la microalgue (séquençage métagénomique) et ont empêché l'acquisition de ce génome. Il a alors été décidé de se concentrer sur ces bactéries pour en apprendre plus sur les interactions entre les populations vivant dans les cultures de *H. ostrearia*. Le sujet de la thèse est donc devenu "**la caractérisation des bactéries retrouvées dans les cultures de *Haslea ostrearia***".

Le chapitre 1 décrit les efforts réalisés pour reconstituer le génome de la microalgue en décontaminant les séquençages ADN. Les informations extraites de ces données ont été les génomes complets ou quasi-complets de la mitochondrie et du chloroplaste, ainsi que de

nombreuses petites séquences du génome nucléaire de *H. ostrearia*. Devant la faible proportion d'ADN de la microalgue par rapport à ceux bactériens, les logiciels d'assemblage ont eu beaucoup de mal à assembler ce génome nucléaire correctement et n'ont donc pas fourni des séquences de qualité suffisante pour être étudiées, ce qui n'a pas été le cas des génomes bactériens retrouvés en grand nombre et entiers, utilisés pour la suite.

Le chapitre 2 et 3 ont été consacrés à la description de ces bactéries. Le deuxième chapitre se concentra sur l'identification des organismes procaryotes présents dans les cultures de *H. ostrearia* sous différentes conditions afin d'évaluer l'effet d'un traitement antibiotique sur ces derniers. L'objectif était de vérifier la reproductibilité du protocole de culture actuellement utilisé, en étudiant la variation de concentration des différentes populations, tout en proposant une première identification de ces bactéries. Les données récupérées pour cela ont été issues d'un séquençage 16S, traditionnellement utilisé pour ce type de problématique. Une trentaine de bactéries marines différentes ont ainsi pu être identifiées. Devant l'observation de populations bactériennes diverses même avec un traitement antibiotique, il a alors été proposé de rechercher des gènes pouvant conférer une résistance à ces derniers durant le troisième chapitre. Les génomes précédemment assemblés ont donc été soumis à une annotation des gènes d'antibiorésistance (ARGs) et de résistance aux antimicrobiens (AMRs). Cette caractérisation par bioinformatique des insensibilités et résistances à ce type de molécules a déjà été plébiscitée par le passé par rapport aux méthodes classiques (antibiogrammes) (Palmer et Kishony 2013, Enne et Bennett 2010, McArthur et Wright 2015). Cette analyse a été également couplée à une vérification de la date d'apparition de ces gènes dans les génomes, qui semble être lointaine et non liée aux récentes cultures de la microalgue en laboratoire. Des recherches complémentaires ont également été menées sur les génomes plasmidiens connus pour faciliter la transmission horizontale d'information génétique, et donc l'acquisition de nouvelles résistances. Au final, une liste de nombreux ARGs et AMRs a pu être dressée et associée à une famille bactérienne.

3. Conclusion et perspectives

À travers l'ensemble des travaux réalisés lors de la thèse, il a été mis en évidence que le but de reconstruction du génome de *H. ostrearia* nécessitait une amélioration de son protocole de culture et/ou d'extraction de l'ADN pour être atteint. En effet, les données ADN provenaient majoritairement d'organismes bactériens, empêchant ainsi la récupération des informations génétiques appartenant à la microalgue. Cette grande diversité bactérienne observée était la raison derrière le fait que le génome de *H. ostrearia* a été "noyé" dans les données de séquençage, et n'a donc pas été correctement assemblé par la suite. Seuls les génomes de ses organelles ont pu être récupérés. Cependant, de nombreux génomes de bactéries ont pu être assemblés en entier et étudiés lors de la thèse. Il a ainsi été montré qu'un traitement antibiotique semblait nécessaire pour la reproductibilité des cultures, mais que ce dernier devait également être choisi avec soin. En effet, de nombreux gènes codant pour des résistances aux antibiotiques ou à des substances antimicrobiennes ont été retrouvés au cours de mes travaux, ainsi que dans certains plasmides.

Ces dernières informations pourraient être utilisées pour observer l'effet de l'absence ou de la présence de certaines bactéries sur la croissance de *H. ostrearia*. Il serait tout à fait possible de croiser les résultats obtenus sous différents antibiotiques donnés pour obtenir les bactéries associées à une augmentation de la croissance de la microalgue. Ces dernières seraient alors considérées comme potentiellement nécessaires pour celle-ci, et conservées dans les cultures. Ceci devrait aussi réduire la diversité bactérienne observée et permettre une meilleure accessibilité du génome de *H. ostrearia*. Des tests avec des antibiotiques pour lesquels aucune résistance n'a été retrouvée dans les génomes bactériens pourront également être réalisés, nous informant ainsi sur la nécessité ou non des bactéries pour la survie de *H. ostrearia*. Ces derniers pourraient également être importants pour comprendre les conditions à réunir pour la production de marennine: ce pigment n'est-il produit que lorsque des bactéries sont présentes dans l'environnement ?

Si des bactéries sont détectées comme essentielles à la survie de *H. ostrearia*, il serait alors intéressant de rechercher ces dernières dans d'autres milieux où a été retrouvée la microalgue. Ces analyses permettraient de comprendre si le microbiome varie en fonction de la localisation de *H. ostrearia*, des conditions intrinsèques à son environnement et les besoins qui en découlent.

Sur un plus long terme, lorsque le génome de *H. ostrearia* sera récupéré et reconstruit, des analyses poussées de ses gènes pourront être effectuées. Il serait alors intéressant de vérifier que les gènes nécessaires à la croissance ou à la vie de la microalgue sont tous retrouvés dans son génome. Et, le cas échéant, si ces derniers sont détectés dans les bactéries précédemment qualifiées d'essentielles pour celle-ci. Les voies métaboliques impactées par chaque organisme pourront également être étudiées dans cette optique.

Introduction

Partie 1 : la diatomée *Haslea ostrearia*

Les diatomées, aussi nommées *Bacillariophyta* (*Ochrophyta*), sont des organismes eucaryotes unicellulaires capables de photosynthèse et vivant dans des milieux aquatiques divers tels que les zones tropicales ou polaires, eaux salines ou douces (Vilmi et al 2015, Malviya et 2016, Otte et al 2023). Ces microalgues sont capables de peupler le fond des eaux (espèces benthiques) ou de vivre en suspension dans l'eau (espèces planctoniques) (Hu et al 2022).

Très largement étudiés, ces organismes ont un rôle écologique important puisqu'ils sont responsables d'environ 40% de la production primaire d'oxygène des océans, soit presque 25% de la production planétaire, tout en ne représentant que 1% de la biomasse photosynthétique terrestre (Marshall et Orr 1928, Moore et al 2001, Fu et al 2022). A cela s'ajoute leur impact majeur sur plusieurs cycles géochimiques (Carbone, Azote, Silice, Phosphate, Fer, etc.) des réseaux trophiques des environnements qu'ils occupent (Fry et Wainright 1991, Allen et al 2006, Ragueneau et al 2006, Sun et al 2011, Muggli et al. 1996). Au vu de leur diversité et de leurs capacités, de nombreuses utilisations en sont faites par l'être humain notamment au niveau industriel (le sédiment diatomite, Zahajská 2020) ou pharmaceutique (Nieri et al 2023). Mais les diatomées sont plus majoritairement utilisées pour évaluer la qualité de l'eau (Indice Biologique Diatomées, norme AFNOR NF T90-354) ou en agroalimentaire (apport nutritif pour l'aquaculture, Lebeau et Robert 2003). D'autres applications sont également potentiellement possibles comme la production de biocarburants (Cadoret et Bernard 2008) ou la phytoremédiation (Retnaningdyah et al 2017).

1. Morphologie des diatomées

Les diatomées sont des algues microscopiques de l'infra-règne des *Heterokonta* (straménopiles), dont la taille des cellules varie entre 10 et 200µm. Mais, bien que les *Heterokonta* ont souvent été caractérisés par la présence de deux flagelles de taille différente, les diatomées n'en possèdent que durant leur reproduction lorsque les cellules sont immatures (gamètes mâles avant la phase végétative, Nanjappa et al 2017, Ferrante et al 2019). Ces microalgues sont entourées d'une paroi cellulaire faite de silice, un exosquelette appelé frustule.

La symétrie et la forme du frustule sont utilisées pour regrouper les diatomées; elles seront appelées pennées si celles-ci possèdent une symétrie bilatérale et une forme allongée, ou centriques si elles sont rondes avec une symétrie radiale (Figure 0-1, Hamm et al 2003). Comme le suggère leur nom en grec ancien, les *diatomos* sont des microorganismes qui ont la particularité d'être "coupés en deux". Leur paroi cellulaire est constituée de deux morceaux de "coque" en silice, deux thèques, qui emboîtées forment le frustule (Figure 0-1). Ces deux thèques/valves légèrement asymétriques, épivalve pour la plus grande et hypovalve pour la plus petite, sont reliées entre elles par des bandes intercalaires nommées *cingulum* (*girdle band*, Figure 0-2) de matière organique. Des multiples motifs et formes du frustule sont ainsi possibles, et caractérisent chaque espèce de diatomées. C'est pourquoi leur identification repose sur les variations observées dans leurs pores, stries et autres éléments morphologiques (Figure 0-1). Par exemple, certaines diatomées pennées peuvent posséder un raphé qui leur octroie la capacité de se déplacer le long d'un substrat, ce qui n'est pas le cas pour d'autres.

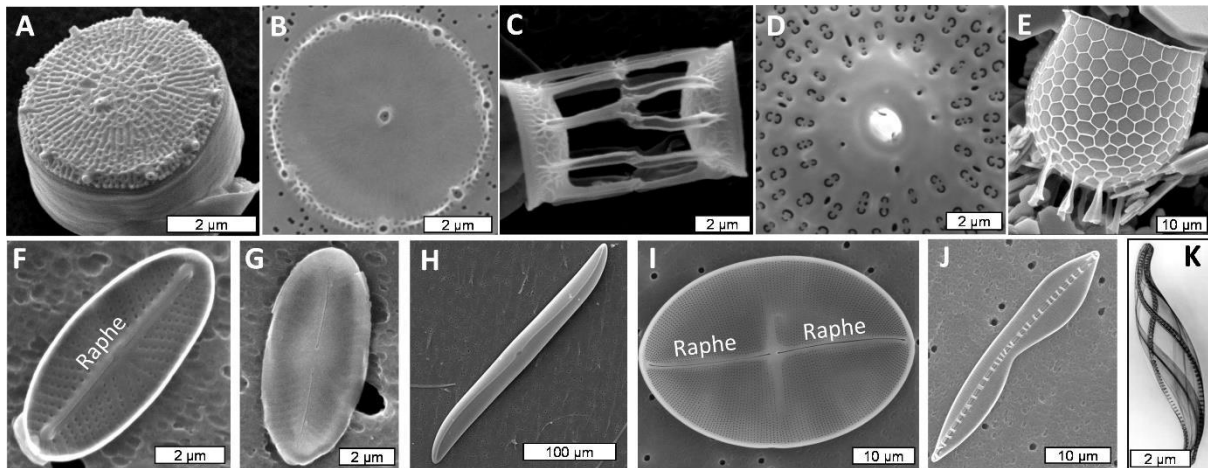


Figure 0-1 : Exemples des structures de valves retrouvées chez différentes espèces de diatomées à symétrie radiale et pennée, Figure extraite de Hildebrand et Lerch (2015) (Reproduced with permission from Elsevier)

(A) *Thalassiosira pseudonana*, (B) *Thalassiosira oceanica*, (C) *Skeletonema costatum*, (D) *Ditylum brightwellii*, (E) *Stephanopyxis turris*, (F) *Navicula pelliculosa* (face proximale), (G) *N. pelliculosa* (face distale), (H) *Gyrosigma balticum*, (I) *Cocconeis* sp., (J) *Bacillaria paxillifer*, (K) *Cylindrotheca fusiformis*. Les diatomées (A) à (E) présentent une symétrie radiale, contrairement à celles (F) à (K) qui sont des pennées. Le raphé est légendé sur les images (F) et (I).

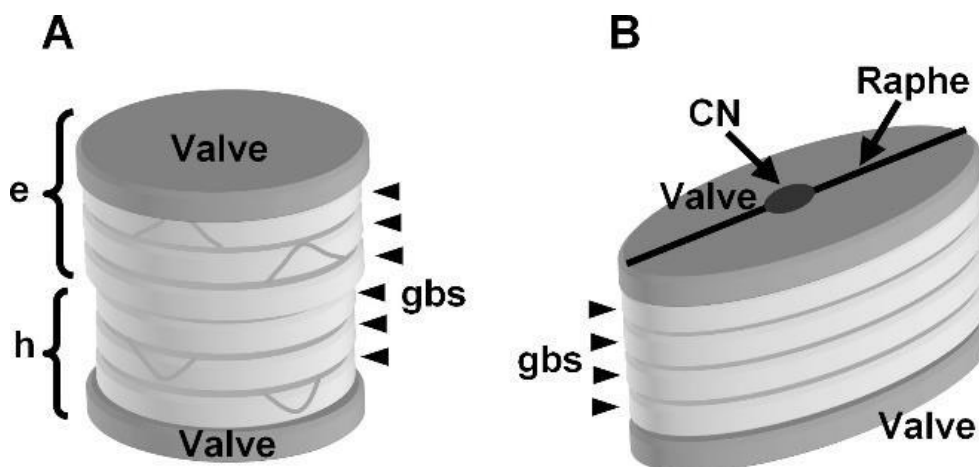


Figure 0-2 : Représentation schématique de la paroi cellulaire des diatomées, Figure extraite de Hildebrand et Lerch (2015) (Reproduced with permission from Elsevier)

Il est ici schématisé la composition de l'épithèque (e) et de l'hypothèque (h) des diatomées centriques (A). Ces thèques comportent chacun une valve et des bandes intercalaires leur permettant le recouvrement de ces derniers. Chez les diatomées pennées (B), on dénote une organisation équivalente mais avec la possible présence d'un raphé.

Le frustule présente également des avantages écologiques, pour lutter contre la prédation, par exemple, ou pour résister à la pression, la silice possédant une résistance plus forte que les exosquelettes composés de calcaire ou de cellulose (coccolithophoridés et dinoflagellés, Hamm et al 2003, Kooistra et al 2007). Le frustule permettrait également de se protéger contre des facteurs abiotiques tels que les UV-B (Ellegaard et al 2016).

Enfin, en faisant varier la densité de son frustule et de ses vacuoles, chaque diatomée est capable d'adapter sa flottabilité. En fonction de leur besoin, ces dernières sont ainsi capables d'adapter leur positionnement dans la colonne d'eau. Ceci a conduit à émettre l'hypothèse selon laquelle des diatomées contaminées ou parasitées seraient en mesure de s'éloigner du reste de la population pour limiter la propagation des maladies (Raven et Waite 2004).

2. Origines évolutives des diatomées

Issues du groupe des *Heterokonta*, les diatomées sont des organismes eucaryotes photosynthétiques possédant des métabolismes très diversifiés : elles peuvent être autotrophes ou hétérotrophes, mais nécessitent des conditions environnementales spécifiques pour survivre. Ceci étant, l'origine évolutive des diatomées est sujette à de nombreux débats, depuis de nombreuses années, et ne semble toujours pas avoir abouti à un consensus à l'heure actuelle.

Selon les estimations basées sur les horloges moléculaires, l'apparition des diatomées date de la jonction entre le Trias et le Jurassique, deux subdivisions de l'ère Mésozoïque, soit il y a environ 200 millions d'années (Kooistra et Medlin 1996). Cette estimation est proche des plus anciens fossiles de diatomées retrouvés à ce jour, puisqu'ils seraient datés d'environ - 190 à -175 millions d'années (Ma) (Rothpletz 1896 et 1900), appartenant au genre *Pyxidicula*. Une grande diversité de diatomées aurait colonisé la planète jusqu'à un important déclin de

leurs populations 160 Ma plus tard, pendant la grande coupure Eocène-Oligocène (-33,9 Ma) (Rabosky et Sorhannus 2009, Lewitus et al 2018). Cet événement semble lié à une compétition interspécifique due à des conditions océaniques de plus en plus difficiles (Iglesias-Rodríguez et al 2002). La diversité des diatomées observées actuellement ne représenterait qu'une infime proportion de celle du Paléogène.

Cependant, l'appartenance des premiers fossiles à ce groupe serait aujourd'hui remise en cause. Brylka et al (2023), par exemple, rejettent l'identification taxonomique de microfossiles issus de trois sites de fouille différents. Initialement associés aux diatomées, ces derniers ne valideraient pas les critères d'évaluation proposés et décrits par cet article : morphologie, contamination, âge, etc. L'étude va même plus loin en réfutant l'appartenance des fossiles de *Pyxidicula* à l'embranchement des diatomées et l'associe à un nanofossile calcaire ou à une amibe, selon leur époque d'origine. Brylka et al (2023) indiquent : "Bien que *Pyxidicula* ressemble à certaines diatomées radiales centriques existantes et qu'elle présente des caractéristiques qui pourraient avoir été similaires à ceux des diatomées ancestrales, nous décrivons de nombreuses sources d'incertitude concernant la fiabilité de ces enregistrements" ("*Although Pyxidicula resembles some extant radial centric diatoms and has character states that may have been similar to those of ancestral diatoms, we describe numerous sources of uncertainty regarding the reliability of these records*", Brylka et al 2023).

L'acquisition de leurs organelles fait également débat. Les diatomées possèdent des chloroplastes bruns-jaunes à quatre membranes, typiques des *Heterokonta*, mais de nombreuses théories tentent d'expliquer leur origine évolutive (Figure 0-3). Toutes prennent leur source après un évènement endosymbiotique primaire, au cours duquel une cyanobactérie ancestrale a été réduite en chloroplaste primaire (Mereschkowsky 1905, Martin et Kowallik 1999).

Selon la théorie des *Chromalveolata*, dans lequel on retrouve les *Heterokonta*, leur ancêtre commun serait un microorganisme doté de chloroplastes et résultant d'un unique évènement d'endosymbiose secondaire : un eucaryote hétérotrophe aurait assimilé et

domestiqué une algue rouge (Cavalier-Smith et al 1999, Yoon et al 2002, Figure 0-3a). Son chloroplaste étant également rouge, il existerait ainsi une étroite liaison entre cet ancêtre et les chloroplastes rouges secondaires retrouvés chez d'autres membres des *Chromalveolata* : *Cryptophyta*, *Haptophyta* et *Alveolata*. Les chloroplastes des diatomées ne seraient que de simples dérivés de ce chloroplaste ancestral (Ben et al 2001).

A cette théorie s'oppose celle d'une origine endosymbiotique tertiaire, l'algue rouge n'ayant pas été directement assimilée par l'eucaryote hétérotrophe, mais *via* l'assimilation d'un autre regroupement d'algues (Figure 0-3b). Ces dernières, supposées être proches des *Cryptophyta* (Dorrell et Bowler 2017, Dorrell et al 2017), auraient domestiqué cette algue rouge avant d'être elles-mêmes englouties par l'ancêtre photosynthétique des *Heterokonta*.

Que l'on regarde une théorie ou l'autre, toutes se rejoignent sur la seule conservation du chloroplaste et de la membrane plasmique des endosymbiotes d'algues eucaryotes au fil du temps. Cette perte des autres compartiments cellulaires aurait formé les chloroplastes de diatomées actuelles, enveloppés dans quatre membranes (McFadden 2001, Figure 0-3c).

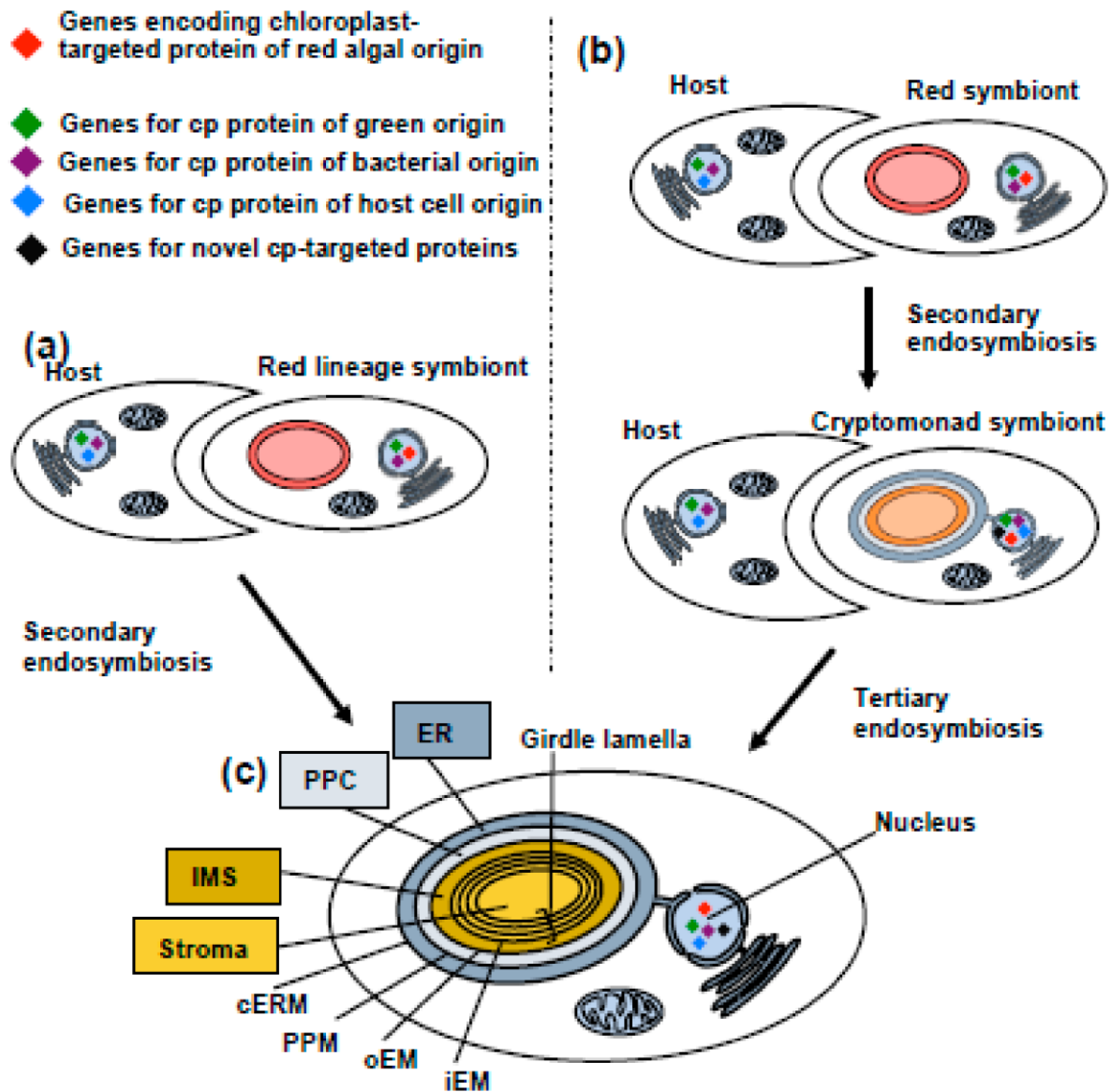


Figure 0-3 : Schéma des deux hypothèses alternatives sur l'origine du chloroplaste des diatomées et de sa structure, Figure extraite de Nonoyama et al (2019)

“(a) Endosymbiose secondaire d'une algue rouge par un ancêtre commun des straménopiles photosynthétiques ou (b) endosymbiose tertiaire d'un organisme de type cryptomonade, hébergeant lui-même un chloroplaste d'origine endosymbiotique secondaire d'une algue rouge. cERM ; membrane réticulaire endoplasmique du chloroplaste ; ER, réticulum endoplasmique ; iEM, membrane de l'enveloppe interne ; IMS, espace intermembranaire ; oEM, membrane de l'enveloppe externe ; PPC, compartiment périplastidique ; PPM, membrane périplastidique.”, description traduite de Nonoyama et al (2019).

3. Classification taxonomique des diatomées

A l'instar de leur origine évolutive très largement débattue, la classification taxonomique des diatomées soulève de nombreuses incertitudes et controverses. De nombreux critères sont apparus au fil du temps pour les regrouper (Figure 0-4).

Gordon et Drum (1994) ont déjà estimé dans les années 1990 l'existence de plusieurs dizaines de milliers d'espèces de diatomées différentes. Ce même nombre a été augmenté à 200.000 espèces quelques années plus tard avec Mann et Droop (1996), longtemps considérée comme l'estimation la plus vraisemblable. Aujourd'hui, les dernières études parlent d'environ 100.000 espèces, dont seulement 10% seraient actuellement décrites (Mann et Vanormelingen 2013, Medlin 2018). Malgré cette estimation revue à la baisse, les diatomées seraient toujours le taxon d'algues possédant la plus grande biodiversité.

Leur grande variété morphologique (taille, forme, etc) observée à travers le temps serait en cause dans le renouvellement systématique des estimations, le concept d'espèce chez ce phylum étant difficile à définir (Partie 1-1, Figure 0-4), et ce même en utilisant différentes définitions pour la notion d'espèce (Mayr 1999, De Queiroz 2007).

Au XVIIe siècle, un débat s'anime autour de la classification des diatomées pour la première fois : doit-on classer ces dernières comme des plantes ou comme des animaux ? C'est ainsi que, puisque les plantes n'ont pas la capacité de se déplacer, les diatomées mobiles ont été classifiées comme appartenant au règne animal, tandis que les formes immobiles ont été associées au règne végétal. Ce n'est qu'au XIXe siècle que l'ensemble des diatomées ont été classées comme "algues", leur activité photosynthétique ayant été observée et validée par les biologistes de l'époque grâce à la démocratisation du microscope (Kützing 1844).

Mettant ainsi fin aux désaccords des membres de la communauté scientifique, une méthode de classification des diatomées a fait son apparition des années après, en 1990

(Round et al 1990, Medlin et al 1996). Basée exclusivement sur les caractéristiques morphologiques générales du frustule des diatomées, cette méthode est devenue celle traditionnellement utilisée par la suite et est définie comme suit : la super-classe des diatomées centriques avec la classe des *Coscinodiscophyceae*, et celle des pennées, divisée en deux classes *Fragilariophyceae* et *Bacillariophyceae*. Cependant, les formes des diatomées étant aussi variées que difficiles à analyser avec les moyens de l'époque, leur classification pouvait s'approcher plus d'une décision arbitraire, que d'une étude basée sur de réelles observations. Les genres et familles ainsi créés ont fait, de nouveau, couler beaucoup d'encre quant à leur véracité. Mais la technologie évoluant à travers les âges, il était alors devenu facile d'observer la structure du frustule en détails, ce qui a rendu la classification de plus en plus précise et a permis l'obtention d'un consensus. De nombreuses espèces ont alors été reclassées, suite à de nouvelles découvertes concernant leurs traits morphologiques (*Fragilariopsis cylindrus* par Mann en 1978 par exemple).

Ce n'est qu'avec l'apparition du séquençage de l'ARN ribosomique 18s (ARNr 18S ou ARN 18S), dans les années 1990, que les premiers arbres phylogénétiques ont été construits pour classer les diatomées. Grâce à ces nouvelles informations moléculaires, de nouvelles anomalies ont été identifiées et corrigées (exemples : Kermarrec et al 2011, Ruck et al 2016). Des révisions majeures se sont avérées nécessaires concernant le système de classification de 1990, très largement utilisé, mais aucun nouveau système n'a réellement été adopté. Il a, par exemple, été proposé par Medlin et Kaczmarska (2004) de former une nouvelle classe, les *Mediophyceae*, à partir de certains ordres de diatomées centriques (Figure 0-4a). Par cette hypothèse, appelée CMB pour *Coscinodiscophyceae+Mediophyceae+Bacillariophyceae*, ces derniers étaient considérés comme intimement liés aux diatomées pennées, et non à celles centriques restantes dans la classe des *Coscinodiscophyceae*. Mais, bien que celle-ci a été acceptée par de grands noms dans le monde des diatomées tels que Mann ou Adl (Medlin et al 2008, Adl et al 2012), d'autres hypothèses de phylogénie ont vu le jour. C'est notamment le cas de Theriot avec sa proposition de gradation structurelle (SGH), qui réfute le groupe monophylétique des *Mediophyceae* en critiquant l'utilisation du seul gène ARNr 18S (Theriot

et al 2009). Cette conclusion est partagée par d'autres, tels que Guo et al (2015) qui a mis en évidence l'absence de cette classe dans les arbres phylogénétiques construits grâce au gène ARNr 18S. Theriot et al ont même poussé plus loin les analyses en utilisant d'autres gènes nucléaires ou chloroplastiques, mais sans jamais réussir à confirmer l'hypothèse CMB (Theriot et al 2010, Theriot et al 2015). La classe des *Mediophyceae* reste aujourd'hui très controversée, bien qu'acceptée par la majorité, et de nombreuses études tentent de confirmer leur existence (dont une grande partie par les auteurs de l'hypothèse initiale : Moniz et Kaczmarska 2010, Medlin 2016, Medlin et Desdevises 2020).

Encore aujourd'hui la classification des diatomées représente un véritable défi et est très largement débattue. Les avancées technologiques grandissantes et les données toujours plus nombreuses sur les diatomées permettent de régulièrement renouveler leur nomenclature. Pourtant, la base de données DiatomBase de AlgaeBase, le registre mondial des espèces marines, utilise encore aujourd'hui les classes définies par Round et al (1990) mais avec l'ajout des *Mediophyceae* de Medlin et Kaczmarska (2004).

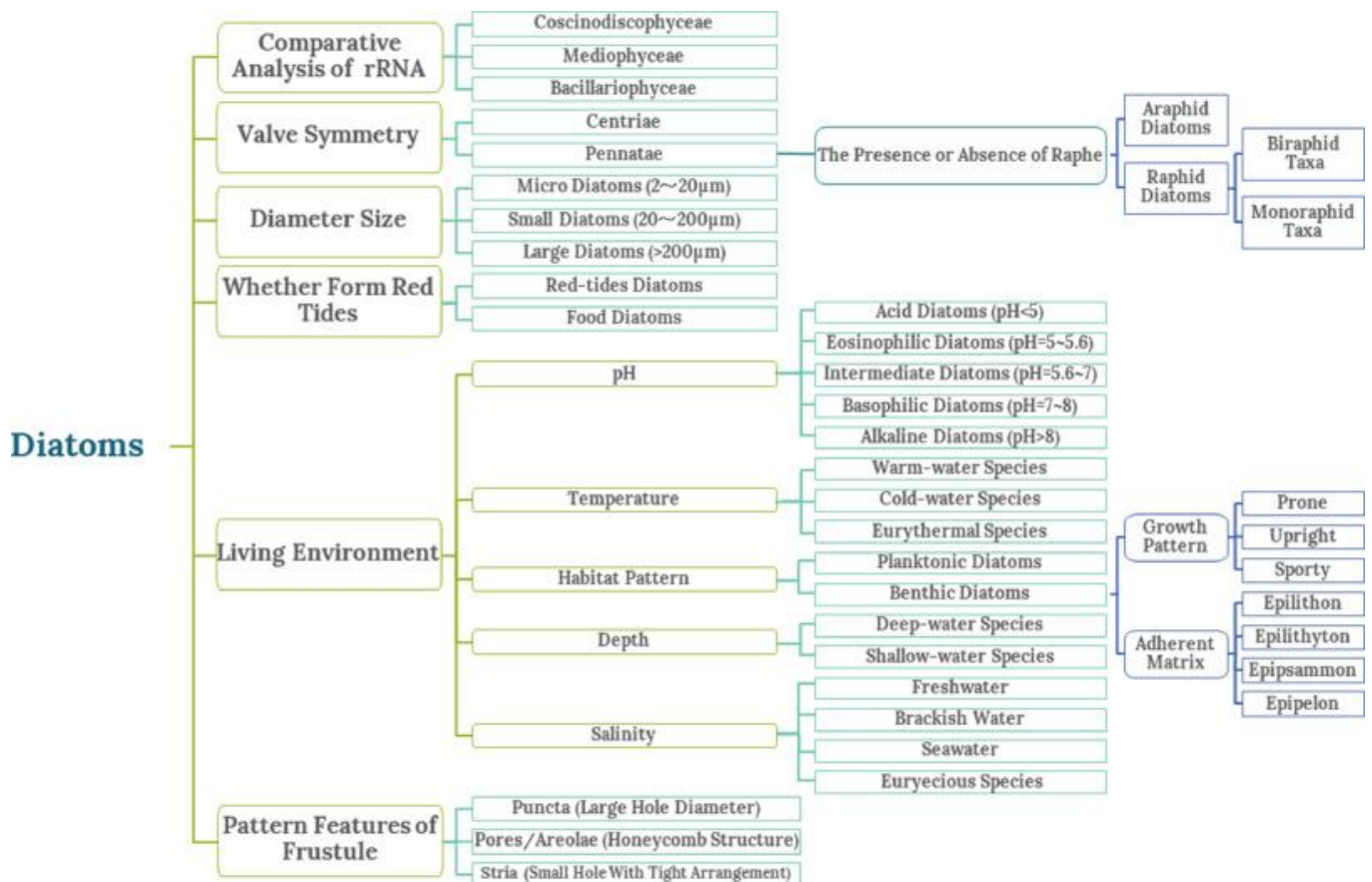


Figure 0-4 : Représentation des différents critères utilisés pour la classification des diatomées, Figure extraite de Fu et al (2022)

Les rectangles au bord arrondis représentent les critères utilisés pour classer les diatomées dans les catégories indiquées par les rectangles.

4. Cycle de vie des diatomées

Les diatomées possèdent deux types de reproduction différentes et peuvent utiliser l'un ou l'autre au cours de leur vie (D'Alelio et al 2009).

La première, et la plus commune, est la reproduction végétative (Figure 0-5a). Étant des organismes unicellulaires diploïdes, la cellule mère est capable de se diviser pour former deux nouvelles cellules filles (mitose). Chacune des deux nouvelles cellules héritent alors de la moitié du patrimoine génétique de la cellule parentale ainsi que d'une de ses deux thèques. Il leur incombe, par la suite, de doubler leur matériel génétique et de synthétiser une deuxième

valve en se servant de celle reçue comme hypovalve (Partie 1-1). Il est à noter que, comme l'une d'entre elles aura obtenu l'épivalve de la cellule mère comme hypovalve, celle-ci sera donc de taille plus petite que l'autre cellule fille. Cette phase végétative peut durer plusieurs années et produire, dans des conditions optimales, l'équivalent d'une division par jour (Rivkin 1986). La seule limite connue à ce jour concernant la reproduction végétative est la taille minimale qu'une cellule doit avoir sous peine d'être dans l'incapacité de survivre. Cette taille critique est, généralement, inférieure à la moitié de la taille originelle de la cellule pour la plupart des espèces (Macdonald 1869).

La deuxième reproduction est celle dite sexuée (Figure 0-5b). Elle est enclenchée lorsque la taille des cellules se rapproche du seuil critique ($\leq 50\%$ taille originelle), et que les conditions environnementales le permettent (Chepurnov et Mann 1997). Cette reproduction sexuée permet à la cellule de retrouver sa taille originelle et d'être le départ d'une nouvelle génération de cellules. D'autres facteurs biologiques ont plus récemment été découverts comme jouant un rôle dans ce déclenchement, tels que la densité cellulaire ou la concentration de partenaires sexuels (Scalco et al 2014, Chepurnov et Mann 2004). A cela s'ajoutent des facteurs environnementaux spécifiques à chaque espèce, comme c'est le cas pour *Haslea (H.) ostrearia*, qui semble nécessiter une quantité et une qualité d'ondes lumineuses spécifiques pour déclencher la reproduction sexuée (Mouget et al 2009). Il est très rare d'observer cette reproduction en laboratoire: premièrement car les cultures sont monoclonales, et donc contiennent un même type de cellule, et deuxièmement par l'absence des conditions environnementales nécessaires.

Les cultures de diatomées en laboratoire, en phase de reproduction végétative, sont difficiles à maintenir sur le long terme car, même s'il semble possible de forcer la reproduction sexuée dans certains cas (Moeys et al 2016, Moore et al 2017), certaines espèces peuvent présenter une dégénérescence génétique due à une reproduction entre clones apparentés (Chepurnov et al 2004). Il s'avère alors nécessaire d'utiliser d'autres moyens pour conserver des isolats/souches sur de longues périodes.

Ainsi, certaines diatomées peuvent par exemple être cryopréservées (Azote liquide+diméthylsulfoxyde) sans difficulté apparente (McLellan 1989, Stock et al 2018), mais ce n'est pas le cas de toutes. Pour reprendre l'exemple de *H. ostrearia*, il est indispensable de développer un protocole de conservation adapté. En protégeant simplement les cellules avec du glycérol avant la cryopréservation seules 10% des cellules restaient viables, contre 60% lorsque ces dernières sont déshydratées avant d'être plongées dans de l'azote liquide (Tanniou et al 2012). Cette dernière méthode reste utilisée à ce jour par manque d'alternatives, bien qu'elle présente de grandes contaminations bactériennes en plus de son taux de survie assez faible. Mais même avec un protocole de cryopréservation adapté, leur décongélation peut s'avérer ardue et entraîner la mort des cellules. De nouveaux protocoles sont alors nécessaires pour assurer l'obtention d'une population viable (exemple : Buhmann et al 2013 avec *Planothidium frequentissimum*).

Lorsque la cryopréservation s'avère trop complexe ou impossible, une alternative existe alors, la réfrigération à basse température (<10°C). Bien que cette méthode ne soit pas parfaite, elle permet néanmoins de ralentir la croissance des cellules et de les conserver sur long et le moyen terme (Kudoh et Takahashi 1989).

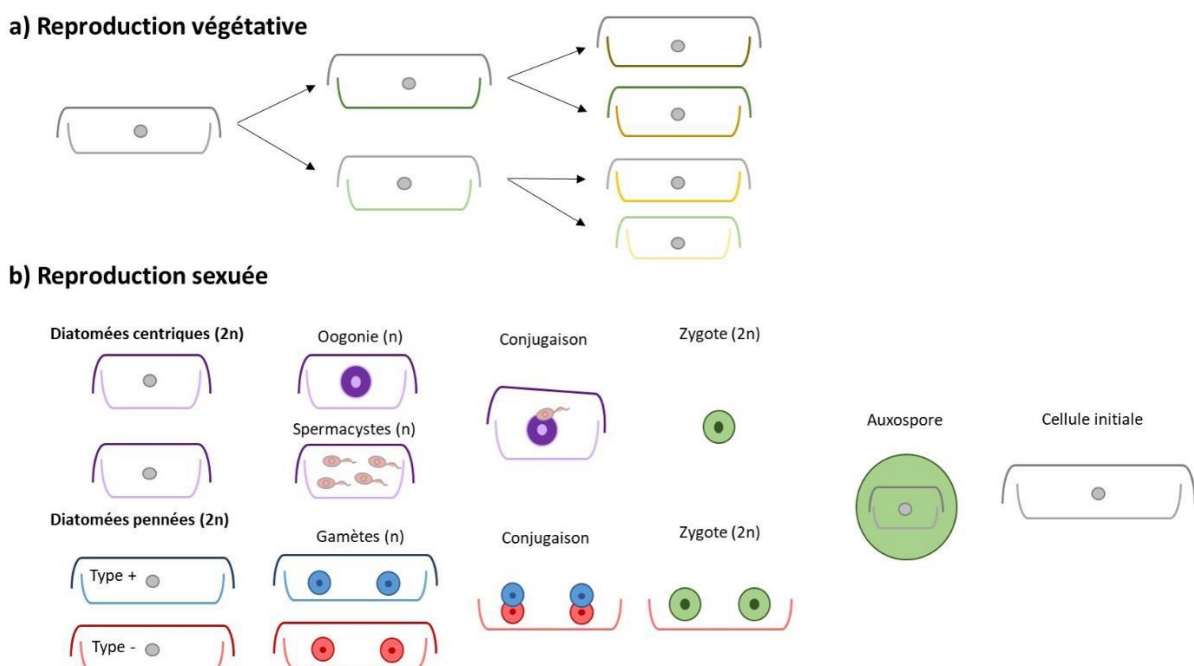


Figure 0-5 : Schématisation des phases de reproduction végétatives (A) et sexuées (B) des diatomées, Figure extraite de Séveno (2022)

5. Interactions biotiques des diatomées

Les diatomées sont au cœur de nombreuses interactions au sein de leur environnement. Il peut s'agir tout autant d'échanges avec d'autres diatomées, d'eucaryotes ou de procaryotes.

Mais avant de nous concentrer sur ce dernier point, voici quelques exemples des nombreuses interactions possibles entre ces microalgues et les eucaryotes environnants.

Premièrement, il existe certaines espèces de diatomées capables de vivre à la surface d'autres animaux ou algues, mais sans pour autant les parasiter (restent à la surface). Il s'agit des diatomées épizoïques ou épiphytes (Tiffany 2011, Letáková et al 2018, Wijewardene et al 2022). Les hôtes sont variés et de toute taille, cela peut aller de tortues (Wetzel et al 2011) aux cétacés (Holmes 1985), en passant par les crustacés (Purushothaman et al 2021), les ectoproctes/bryozoaires (Asagabaldan et al 2023), les algues (Khan et al 2020) et d'autres plantes aquatiques (Prazukin et al 2022). La plupart de ces interactions sont inoffensives pour l'hôte, les diatomées épiphytes se nourrissant de silicates, de phosphates, de nitrates ou de composés organiques dissous dans l'environnement.

Deuxièmement, les diatomées sont également capables d'allélopathie. Cette dernière se définit par l'ensemble des interactions biochimiques réalisées par les plantes entre elles (Molisch 1937), et est très étudiée par la communauté scientifique dans le cas des diatomées planctoniques. Celles-ci sont capables de produire un grand nombre de composés allélopathiques (Acides Gras PolyInsaturés, Aldéhydes PolyInsaturés, etc.) qui peuvent affecter négativement ou positivement la croissance des autres espèces concurrentes du phytoplancton présentes dans l'environnement (Ali et al 2021). L'objectif est de maintenir leur domination dans le milieu en stimulant ou en excréant des produits chimiques pour modifier le "fonctionnement des communautés, des populations et des écosystèmes, y compris la défense, la reproduction, la concurrence et l'évitement des prédateurs" ("*functioning of*

communities, populations, and ecosystems including defense, reproduction, competition, and predator avoidance”, Ali et al 2021).

Enfin, pour conclure cette liste (non exhaustive), des études ont démontré que les diatomées peuvent être sujettes au parasitisme par des organismes tels que les *Chytridiomycota* (*Fungi*, Bruning 1991) ou d'autres stramenopiles (*Oomycota*, Thines et al 2015).

De nombreuses interactions entre les diatomées et les procaryotes ont également été remarquées, en milieu naturel comme en laboratoire (Vincent et Bowler 2022). La majorité de ces échanges sont réalisés avec des bactéries, dont on peut distinguer deux catégories : les bactéries épiphytes/du microbiome et celles libres. Les premières vivent à la surface des diatomées, dans la fine couche d'eau qui les entourent (phycosphère), et les deuxièmes se situent en suspension dans la colonne d'eau. Les effets de ces interactions sont vastes et peuvent être positifs, négatifs, ou neutres pour chacun des organismes participants (Figure 0-6).

Ainsi une relation mutualiste peut se former entre une diatomée et une bactérie, relation où chacune en ressortira gagnante. Dans le cas des diatomées autotrophes, les bactéries du microbiome peuvent se nourrir de composés relâchés par la microalgue (carbone organique dissous) (Liu et al 2022). Ces derniers, reminéralisés, pourront être de nouveau exploités par la diatomée. Mais cette relation gagnant-gagnant ne s'arrête pas là. Il n'est pas rare d'observer une recrudescence de la concentration de certaines bactéries lors de l'efflorescence algale (*bloom*) des diatomées (Hou et al 2020, Shao et al 2020). Bien qu'aucune étude ne prouve l'action directe des microalgues sur les bactéries environnantes, les concentrations bactériennes semblent pourtant varier avec le temps et les besoins des diatomées. Il a été de ce fait suggéré que les procaryotes sont sélectionnés par la microalgue selon leurs bienfaits envers elle, comme l'apport de nutriments (Teeling et al 2012, Shibl et al 2020, Sterling et al 2023). Les bactéries sont aidées par la microalgue pour leur développement dans le milieu, et celle-ci les utilise pour accroître sa croissance. Les exemples les mieux documentés sont ceux

de la vitamine B12 et des sidérophores synthétisés par les bactéries. Le premier est un composé indispensable à la croissance de la microalgue (Sultana et al 2023), et l'autre facilite l'accès au fer à cette dernière (Gao et al 2021). Mais, pour s'assurer que cette relation mutualiste se déroule bien, une communication doit s'instaurer entre les deux protagonistes. Celle-ci peut être chimique, comme une étude de Amin et al (2015) le suggère, en concluant que "La capacité des bactéries à détecter les composés animaux et végétaux suggère que les bactéries marines associées aux diatomées devraient être capables de détecter les phéromones sécrétées ou les molécules de signalisation" ("*The ability of bacteria to sense animal and plant compounds suggests that marine bacteria associated with diatoms should be able to sense excreted pheromones or signaling molecules*", Amin et al 2015). Cette étude se base sur une *Sulfitobacter* dont la production d'hormones (acide indole-3-acétique) permet de stimuler la division cellulaire de *Pseudo-nitzschia multiseriis*. Ainsi, la diatomée est en mesure de fournir à la bactérie la taurine qui lui est nécessaire pour sa survie. Dans la bibliographie, on distingue deux grands genres de bactéries décrits comme ayant un effet positif sur la croissance des diatomées : les *Sulfitobacter* et les *Rhodobacteraceae* (Amin et al 2015, Johansson et al 2019).

D'autres relations peuvent également s'établir avec un effet neutre ou négatif pour la diatomée. Pour reprendre l'exemple précédent des bactéries se nourrissant des composés expulsés par la microalgue, il est aussi tout à fait possible que ces dernières se nourrissent des molécules composant la diatomée à sa mort (Waksman et Butler 1937). Il s'agira alors d'une relation de commensalisme, où seule la bactérie est gagnante. De même, des interactions aux effets négatifs pour l'un des deux organismes existent (prédation, compétition, parasitisme, amensalisme). Certaines bactéries sont en effet capables de produire des protéases anti-algales (Kim et al 2005), toxiques pour la microalgue, ou de causer leur lyse grâce à la création de cellules filamenteuses multicellulaires (regroupement par films bactériens, Furusawa et al 2003). A l'inverse, les diatomées sont également en mesure de créer des molécules antibactériennes (Lebeau et Robert 2003). Il s'agira ici d'une relation d'amensalisme. Une compétition indirecte peut également s'établir entre les bactéries et les

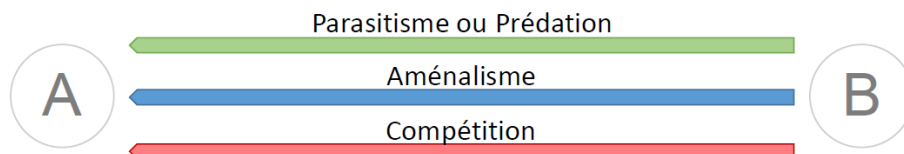
diatomées : les ressources présentes dans le milieu étant limitées (phosphore, azote, etc.), la première à les assimiler prend l'ascendant sur l'autre (Thingstad et al 1993).

De manière générale, et sans qu'il soit fait mention d'effets sur les cultures, les bactéries associées aux diatomées appartiennent majoritairement aux *Bacteroidetes* et aux *Gammaproteobacteria*, puis aux *Alphaproteobacteria* (Sapp et al 2007, Klindworth et al 2014, Jauffrais et al 2017, Sterling et al 2023). Les concentrations semblent cependant varier selon les conditions environnementales, les besoins des diatomées (Sapp et al 2007, Shibl et al 2020, Sterling et al 2023), et selon l'espèce étudiée (Koedooder et al 2019, Helliwell et al 2022).

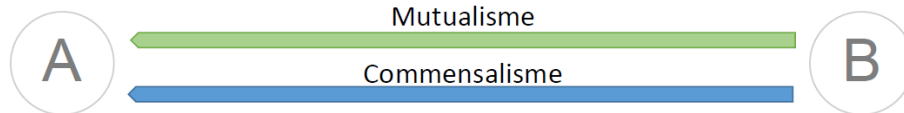
A travers ces nombreux exemples, il est facile de comprendre qu'un équilibre fragile relie les diatomées et les organismes qui les entourent (Tilstone et al 2000, Johansson et al 2019, Shao et al 2020). Ces interactions seraient anciennes puisque plusieurs études ont retrouvé des gènes bactériens dans des génomes de diatomées (Bowler et al 2008, Armbrust 2009). Ce transfert aurait eu lieu suite à la proximité et l'étroite collaboration entre les bactéries et les diatomées au cours de leur évolution. Pour comprendre ces microalgues, il serait donc nécessaire de ne pas les étudier seules mais comme des organismes possédant leur propre microbiome et indissociables de ce dernier (Youle et al 2013, Behringer et al 2018).

Cependant, comme il est d'usage chez ces microalgues, cette considération des diatomées et de leurs bactéries comme un "tout" est sujette à débat. Notamment si l'on considère que les symbiotes bactériens peuvent avoir colonisé d'autres organismes (non exclusifs aux diatomées) et/ou avoir des pressions de sélection différentes de celles de la microalgue (Moran et Sloan 2015, Douglas et Werren 2016).

(A) Lorsque l'espèce A a un effet **néгатif** sur l'espèce B :



(B) Lorsque l'espèce A a un effet **positif** sur l'espèce B :



Légende :





-  Effet positif
-  Effet neutre
-  Effet négatif
-  Sens de l'effet

Figure 0-6 : Schématisation des types d'interactions possibles entre espèces, Figure inspirée de Vincent et Bowler (2022)

6. Les caractéristiques du sous-genre *Haslea* bleue

Les *Haslea* (*H.*) sont un genre de diatomées de la famille des *Naviculaceae*. Elles réunissent des algues microscopiques dont les cellules sont enfermées dans une "coque" de silice en forme de *navicula*, "petit bateau" en français. Il s'agit de microalgues pennées avec un raphé positionné le long de l'axe apical et un sternum étroit, ce qui donne aux valves un aspect cruciforme. Comme toutes les diatomées, ces dernières sont capables de synthétiser de nombreux pigments classiques du phylum (Mouget et al 1999), mais certaines d'entre elles se distinguent des autres diatomées. En effet, leurs cellules possèdent une couleur bleutée issue d'un pigment de la même couleur : la marennine (Prasetya et al 2020, Simonsen 1974, Figure 0-7a). Actuellement 5 espèces différentes composent le sous-genre des *Haslea* bleues : *H. ostrearia*, *H. karadagensis*, *H. provincialis*, *H. nusantara* et *H. silbo*.

La marennine est connue depuis 1886 (Moreau 1967) mais encore mal caractérisée à ce jour, malgré les études menées par quelques laboratoires, dont le nôtre. Ce pigment pourrait en effet avoir de nombreuses applications en agroalimentaire et cosmétique. L'un des principaux objectifs serait, à terme, de remplacer des colorants pétrochimiques tels que l'indigo carmine E132 ou le bleu brillant FCF E133 par ce pigment naturel (Sigurdson et al 2017).

D'autres propriétés, comme un effet antiviral et antibactérien pourraient être intéressantes en aquaculture (Latour et al 2023, Bhattacharjya et al 2023) et pourraient être aussi évaluées d'un point de vue biomédical (Gastineau et al 2012, Falaise et al 2019). Malheureusement, pour ce dernier effet, ces propriétés n'ont été observées que dans un petit nombre d'études en laboratoire et restent donc au stade d'hypothèse. L'une des principales études sur le sujet a été réalisée en 2012 et a démontré une inhibition de la croissance de 3 souches de bactéries marines par la marennine à 1µm/mL, mais n'a permis de prouver l'efficacité du pigment qu'à fortes doses sur le virus de l'herpès et un manque d'effet sur 9 souches de bactéries pathogènes terrestres (Gastineau et al 2012). Les deux "formes" utilisées du pigment ont été récupérées en filtrant le substrat et le surnageant des cultures de *H. ostrearia*, comme proposée par Pouvreau (2006). Ainsi il n'est pas prouvé que la marennine en soit le seul composant, comme le met en avant Bélanger et al (2020). De plus, dans cette étude, les auteurs ne font état d'aucun contrôle négatif pour les tests sur l'activité antibactérienne (tests réalisés sans marennine). D'autres études ont été tentées à la suite de ces travaux, mais l'effet antibactérien n'a majoritairement été testé que sur des espèces de *Vibrio* et avec le substrat des cultures de *H. ostrearia* (Falaise et al 2019). Enfin, la marennine semblerait réguler "les communications intercellulaires entre les astrocytes et les cellules microgliales" dans un contexte d'immunomodulation ("*These findings suggest an interesting role of marennine in immunomodulation by regulating intercellular communications between astrocytes and microglial cells*", Méresse et al 2023), ainsi que posséder des activités anti-inflammatoires et anticoagulantes qui lui apporteraient, là encore, un intérêt clinique (Ale et Meyer 2013).

Si la composition chimique exacte du pigment n'a pas encore été déterminée, la marennine serait une molécule complexe et les dernières hypothèses lui associent un "squelette carboné composé de polysaccharides, de groupements polyphénoliques et un chromophore" (Zebiri et al 2023, Pouvreau et al 2006). Son origine est également encore inconnue, les mécanismes entrant en jeu pour la synthèse du pigment n'ayant pas encore été décrits. Seules des hypothèses existent quant à son intérêt pour la cellule, vis-à-vis des putatives propriétés mentionnées précédemment qui lui conféreraient un avantage face aux

autres organismes de son environnement. Les connaissances actuelles se résument au lieu de stockage de la marennine, qui se situe dans le cytoplasme des cellules d'*Haslea*, aux extrémités (Figure 0-7a), et aux deux formes extracellulaires et intracellulaires que ce pigment peut revêtir (Pouvreau et al 2006).

L'espèce emblématique du sous-genre bleu est *H. ostrearia* (Gaillon) Simonsen (Gaillon 1820, Simonsen 1974). Elle a été découverte pour la première fois en 1820 par Gaillon lors de recherches sur un phénomène jugé étrange à l'époque : le verdissement des branchies d'huîtres dans plusieurs parcs ostréicoles français d'Atlantique et de la Manche (Marennnes d'Oléron, Havre et Dieppe) (Gaillon 1820, Figure 0-7b). En automne, il est observé des "masses d'un vert noirâtre" tapissant le fond des claires ostréicoles (affinage, 0 à 15m de profondeur, Figure 0-7c). De ces amas récupérés et analysés au microscope, Gaillon y a observé des cellules d'une centaine de micromètres de long qu'il nomma *Vibrio ostrearius*. Il a fallu attendre 1974 pour qu'elles soient rebaptisées *H. ostrearia* (Gaillon) Simonsen par Simonsen, lors de leur redécouverte durant une expédition dans l'océan indien (Simonsen 1974). Grâce aux connaissances plus approfondies des diatomées au XXe siècle et l'invention de la microscopie électronique à balayage, l'espèce appartient alors à son propre genre, *Haslea* : en hommage à la spécialiste du phytoplancton G.R. Hasle. D'autres microalgues y sont également transférées, dont quelques espèces de *Amphipleura*, *Navicula*, *Gyrosigma* et *Pleurosigma*. Aujourd'hui les huîtres colorées par la fixation de la marennine sur les branchies ne sont plus jugées comme étranges, et représentent même un intérêt économique pour les ostréiculteurs par leur coût plus élevé que les huîtres "normales".

H. ostrearia est retrouvée dans de multiples environnements (tempérés à tropicaux, Figure 0-8) et est au cœur de nombreuses études. Par exemple, la majorité des études réalisées pour la compréhension du pigment de la marennine sont réalisées sur cette espèce (Pouvreau 2006, Gastineau et al 2012, Falaise et al 2019, Méresse et al 2023), tandis que d'autres préfèrent se concentrer sur le fonctionnement de la microalgue, notamment en étudiant l'effet sur les cultures des composants protéiques du potentiel microbiome de cette

dernière (Lépinay et al 2016, 2018). Pour autant l'espèce reste assez méconnue à ce jour, bien que des programmes de recherche s'articulent autour d'elle tels que le projet européen H2020 GHANA ou le pari scientifique LANCOM dont est issue cette thèse.



Figure 0-7 : Vue microscopie de *Haslea ostrearia* (A), vue d'une huître verdie par la présence de marennine sur ses branchies (B) et vue d'un bloom de *Haslea ostrearia* (C)

Les images ont été extraites de la fiche descriptive de *H. ostrearia* disponible sur le site des Données d'Observations pour la Reconnaissance et l'Identification de la faune et la flore Subaquatiques (<https://doris.ffessm.fr/Especies/Haslea-ostrearia-Navicule-bleue-5595>). La photo (C) a été prise à Calvi à 4 m de profondeur, et la photo (B) représente une huître Claire de la Tremblade.



Figure 0-8 : Localisation des *Haslea ostrearia* à travers le monde d'après la littérature, Figure extraite de Séveno (2022)

Les informations données ici n'ont pas été vérifiées par données moléculaires, certaines ne sont issues que d'observations visuelles ou microscopiques.

7. La génomique de *Haslea ostrearia*

Comme pour de nombreuses diatomées, les connaissances en génomique sur *H. ostrearia* sont limitées. Pour l'ensemble des diatomées, seules 7 publications font état d'un génome reconstitué en entier, *i.e.* assemblé jusqu'au niveau chromosomique (Tableau 0-1), et ce malgré de nombreuses tentatives sur différentes espèces. A travers les années, il a été observé l'accumulation de données partielles pour cet embranchement de microalgues. Plus d'une soixantaine de génomes ont été séquencés depuis la fin des années 2000 (Tirichine et al 2017, selon le nombre de génomes associés aux *Bacillariophyta* sur le NCBI le 10/12/2023), mais ces efforts ont produit des assemblages constitués de plusieurs centaines de séquences et ne sont donc pas considérés ici comme suffisamment aboutis pour être qualifiés d'« entiers » mais plutôt comme des génomes « partiels » (*draft genome*) tel que le très souvent cité génome de *Pseudo-nitzschia multiseries* (JGI 2011). Ce phénomène est notamment dû à l'impossibilité d'obtenir des cultures pures/en conditions d'axénie (sans aucun

autre organisme que la diatomée étudiée). Un récent projet de 2020 vise à augmenter le nombre de génomes séquencés de diatomées à 100 (Mock 2020).

A ce jour, les génomes nucléaires entiers fréquemment cités et utilisés du phylum sont ceux de *Thalassiosira pseudonana* (Armbrust et al 2004) et de *Phaeodactylum tricornutum* (Bowler et al 2008) (Tableau 0-1). Ces espèces sont connues pour être faciles à cultiver en laboratoire et en conditions axéniques, c'est pourquoi leurs génomes ont pu être reconstitués entièrement. Cependant, il est difficile de les utiliser en tant que génome de référence pour les autres diatomées, dont *H. ostrearia*. En effet, comme évoqué précédemment leur taxonomie est en constante évolution au fil des découvertes, et à l'heure actuelle il est donc difficile d'estimer à quel point une diatomée est proche d'une autre génétiquement et/ou évolutivement parlant (Partie 1-3). De plus, ces deux diatomées possèdent des spécificités qui semblent les éloigner de celle pennée benthique et marine qu'*H. ostrearia*. *Thalassiosira pseudonana* est une diatomée centrique qui ne se serait adaptée aux environnements marins que récemment (Alverson et al 2011). *Phaeodactylum tricornutum*, bien qu'elle soit également une microalgue pennée marine, présente une morphologie particulière. Il a été observé des phénomènes singuliers pour le phylum comme des variations de forme de sa paroi cellulaire et une synthèse de silice seulement possible sous sa forme ovale (Desclés 2007). Face à cela, chaque analyse se voit contrainte d'utiliser plusieurs souches différentes pour conclure sur l'espèce, rendant difficile l'utilisation de son seul génome de 2008 comme référence (Martino et al 2007). Les autres génomes partiels de diatomées ne semblent également pas adaptés à une utilisation comme génome de référence au vu de leur nature, et des problèmes taxonomiques soulevés plus tôt. Les seules informations que l'on peut alors tirer de ces données, et des 5 autres génomes entiers connus de diatomées, sont une estimation de la taille du génome d'une diatomée ainsi qu'une tendance en ce qui concerne sa composition (Tableau 0-1). Ainsi, *H. ostrearia* a de grandes chances d'avoir un génome d'au moins 20 millions de bases avec un pourcentage de guanine-cytosine proche de 47% (%GC), les génomes assemblés étant connus pour avoir une taille réduite par rapport aux autres espèces du phylum (Falciatore et

al 2020). Il y a aussi de fortes probabilités de retrouver un pourcentage non négligeable de gènes bactériens à l'instar de *Phaeodactylum tricornutum* (5%, Bowler et al 2008), ainsi que des gènes avec une origine évolutive complexe. Pour prendre *Thalassiosira pseudonana* en exemple, on retrouve tour à tour de fortes similarités entre les séquences de ses gènes avec des homologues provenant d'algues rouges, parfois d'algues vertes mais également d'animaux (Armbrust et al 2004). Dans le cas de ce dernier, on peut notamment citer des gènes codant pour des enzymes prenant part dans le cycle de l'urée, voie clef pour la fixation de composés essentiels à la croissance des diatomées (Allen et al 2011).

Espèce	Souche	Niveau d'assemblage	Taille (pb)	%GC	Nombre de gènes	Dernière mise à jour	Référence
<i>Thalassiosira pseudonana</i>	CCMP1335	Chromosome (23)	32.437.365	46,5	11.771	25/11/2015	Armbrust et al 2004
<i>Skeletonema marinoi</i>	CNS00100	Chromosome (24)	65 Mb	45,5	N/A	08/09/2023	Liu et al 2023
<i>Phaeodactylum tricornutum</i>	CCAP 1055/1	Chromosome (33)	27.450.724	48,5	10.398	12/12/2008	Bowler et al 2008
<i>Nitzschia sp. pyKryTriq1</i>	CATPKY01	Chromosome (17)	40,5 Mb	48	N/A	14/07/2023	NCBI GCA_956465555.1
<i>Nitzschia sp. DOCU1</i>	CATOAN01	Chromosome (17)	52 Mb	49	N/A	13/05/2023	NCBI GCA_951394045.1
<i>Fistulifera solaris</i>	JPCC DA0580	Chromosome (44)	51,2 Mb	46	N/A	01/09/2022	Maeda et al 2022
<i>Epithemia pelagica</i>	CRS-2021b	Chromosome (15)	60,3 Mb	48	N/A	07/04/2023	NCBI GCA_946965045.2

Tableau 0-1 : Tableau récapitulatif des caractéristiques des génomes entiers de diatomées actuellement disponibles

Les informations ont été récupérées depuis les dernières versions mise à jour des génomes disponibles en Décembre 2023.

Plus d'informations sont cependant disponibles en ce qui concerne le génome de ses organelles. D'après ces dernières, chaque cellule comporte deux chloroplastes allongés le

long de leur paroi cellulaire (Bedoshvili et al 2009), accompagnés de plasmides (Karas et al 2015) ainsi que des mitochondries (Prihoda et al 2012). Longtemps cantonné à des séquences partielles pour les gènes *rbcL* provenant du chloroplaste et *COI/cox1* de la mitochondrie, ce n'est que très récemment que ces génomes ont pu être reconstruits en entier pour deux souches différentes (Chapitre 1, Dąbek et al 2022). Comme l'une d'entre elles sera détaillée par la suite, les informations présentées ici se concentreront sur l'isolat provenant de l'océan indien et publié en 2022.

La mitochondrie est constituée d'un chromosome circulaire à simple brin d'une taille de 37 444b avec %GC de 28,88, et possède de grandes similarités avec celles retrouvées chez d'autres algues de couleur brune. D'un total de 62 gènes (38 CDS, 22 ARNt et 2 ARNr), elle contient le code génétique de protéines indispensables car liées à la chaîne respiratoire ou aux sous-unités ribosomiques. On y retrouve également le gène *COI*, utilisé pour la construction d'arbre phylogénétique et l'estimation de la distance génétique entre les différentes espèces de *Haslea* ou de diatomées (Gabed et al 2022). Enfin des divergences ont été observées par rapport au génome mitochondrial d'autres *Haslea*. Par exemple, aucun intron n'est détecté dans son génome "comme celui de *H. nusantara*, ce qui n'est pas le cas de *H. silbo*" ("like *H. nusantara* and unlike *H. silbo*, no intron was detected in the mitogenome", Dąbek et al 2022). Des positions de cadres de lecture (ORFs) changeraient aussi selon l'espèce *Haslea* considérée.

Quant au chloroplaste, mesurant 129 009b pour un taux de GC de 31,18%, il semble typique de ceux assemblés pour les autres diatomées. Son architecture en 4 parties comprend 176 gènes au total (144 CDS, 29 ARNt et 6 ARNr), répartis dans sa région courte à copie unique, sa région longue à copie unique et ses deux régions répétées inversées (Oudot-Le Secq et al 2007, Hamsher et al 2019). Parmi eux, plus de 70 gènes associés à des protéines conservées au sein des diatomées ont été retrouvés, dont le gène *rbcL* utilisé pour l'étude phylogénétique entre les différentes diatomées (Gabed et al 2022). Cette nouvelle séquence complète permet d'enrichir les bases de données actuelles sur les diatomées, jugées insuffisantes par la communauté scientifique (Turk Dermastia et al 2023). Deux plasmides ont

également été retrouvés et ont été associés au chloroplaste, de tailles respectives de 4,343b et 3,313b. Chacun semble contenir des ORFs non conservés par le genre *Haslea*, mais ils partagent de fortes similarités avec ceux annotés pour *H. silbo* comme la présence du gène *serC* sur l'un d'eux.

Partie 2 : L'assemblage de génomes

"Full disclosure: Genome assembly is still hard." – Stephen Richards, *PLoS Biology* 16(4): e2005894 (2018). <https://doi.org/10.1371/journal.pbio.2005894>

1. Histoire du séquençage de l'ADN

Les molécules de *nuclein*, connues aujourd'hui sous le nom d'ADN (Acide DésoxyriboNucléique), ont été observées pour la première fois sous forme de précipité blanchâtre en 1869 (Miescher 1869). Il a fallu attendre les années 1950 pour que des découvertes majeures soient faites quant à sa composition (Dahm 2005). En 1953, un article historique de Watson et Crick sort dans le journal scientifique *Nature*, il y décrit la structure de l'ADN (Watson et Crick 1953). Dix ans plus tard, un prix Nobel est remis à Holley, Khorana, Nirenberg et al pour avoir réussi à déchiffrer le code génétique (Singer 1968). Il est alors connu que les molécules d'ADN peuvent encoder les informations génétiques du vivant avec seulement 4 composés organiques différents, les nucléobases Adénine (A), Thymine (T), Cytosine (C) et Guanine (G), dont chacun est associé à un sucre et de un à trois groupements phosphates pour former un nucléotide.

De ces connaissances s'ensuit, en 1977, la démocratisation d'une des premières techniques de détermination d'une séquence d'ADN : le séquençage Sanger (Sanger et al 1977). Cette nouvelle méthode révolutionnaire a permis de séquencer l'ADN du bactériophage phiX-174, puis celui d'autres microorganismes dont le premier génome eucaryote avec la

levure *Saccharomyces cerevisiae* (Goffeau et al 1996), le premier génome humain en 1996 (90% de la séquence, ou encore, la première diatomée avec *Thalassiosira pseudonana* en 2004 (Armbrust et al 2004). Cependant, malgré sa grande précision, le séquençage Sanger coûte cher et se montre très fastidieux à mettre en place. Seuls les organismes "modèles", très connus sur le plan expérimental et étudiés par une grande partie de la communauté scientifique mondiale, y ont donc accès. On prendra pour exemple le projet "Génome humain" (*Human Genome Project*) de 1988 dont le but était de séquencer entièrement le génome humain. Il aura fallu presque 15 ans (fin en 2003) et 300 millions de dollars au *National Institute of Health* (NIH) pour établir le séquençage quasi-complet de l'ADN du génome humain, et plus de 30 ans pour l'obtenir de télomère à télomère, sans aucun "trou" (*gap*) dans le génome (Nurk et al 2022). Pour des raisons historiques, il est précisé que la méthode Sanger fut publiée parallèlement avec celle de Maxam-Gilbert qui a aidé à révolutionner l'histoire de la biologie moléculaire en séquençant chimiquement l'ADN (Maxam et Gilbert 1980), mais n'est pas détaillée ici car quasiment abandonnée de nos jours.

Ce n'est qu'après l'apparition de nouvelles techniques de séquençage, appelées technologies de séquençage haut-débit (*High Throughput Sequencing*) ou de nouvelle génération (*New or Next Generation Sequencing*), que les génomes d'espèces peu étudiées par la communauté scientifique ont pu être déterminés, comme *Phaeodactylum tricornutum* (Bowler et al 2008) et *Pseudo-nitzschia multiseries* (JGI 2011). C'est ainsi que le coût est passé, pour un génome humain, d'environ 100 millions de dollars en 2001 à moins de 1,000\$ aujourd'hui (Figure 0-9). Les nouvelles techniques de séquençage utilisées de nos jours appartiennent à la seconde et troisième génération.

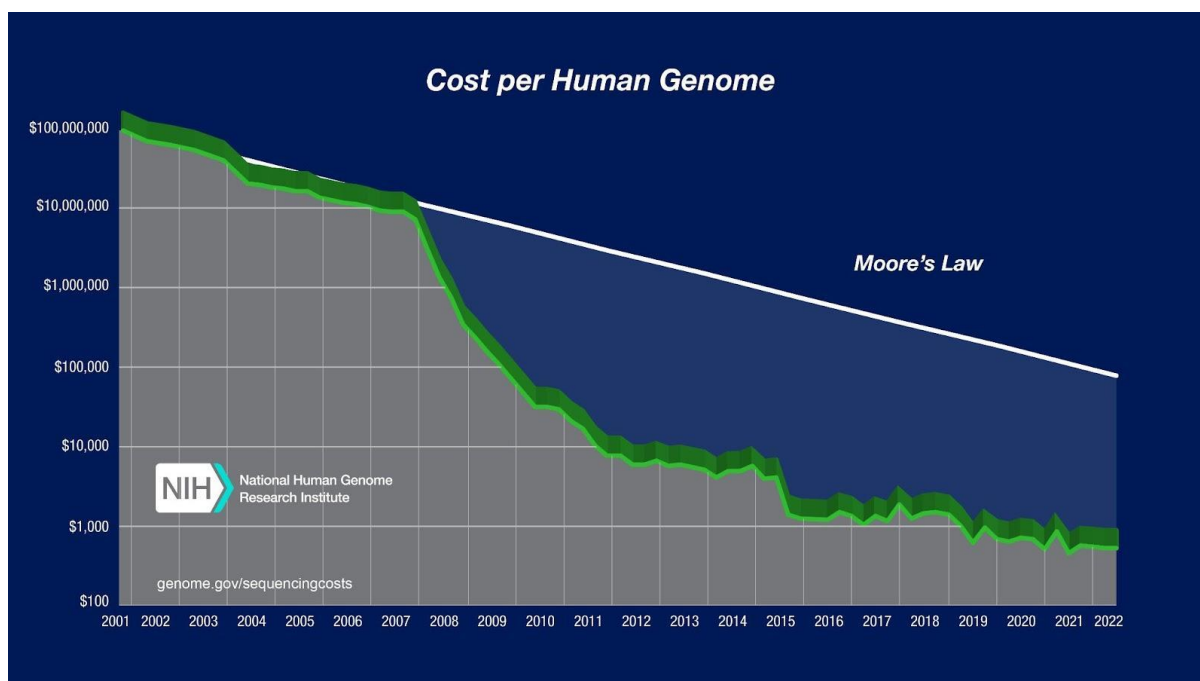


Figure 0-9 : Courbe de l'évolution du prix d'un séquençage du génome humain de 2001 à 2022, Figure extraite du site [genome.gov/sequencingcosts](https://www.genome.gov/sequencingcosts)

Ce graphique représente l'évolution du prix du séquençage du génome humain de 2001 à 2022, et la compare aux prédictions faites par la loi de Moore en 2008. L'image est disponible sur le lien suivant : <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>.

©National Human Genome Research Institute

2. Les techniques de séquençage

Comme expliqué plus tôt, la première technique de séquençage de l'ADN utilisée à grande échelle est celle mise au point par Sanger en 1977 (Tableau 0-2). Dans sa méthode originelle, l'ADN est extrait de l'échantillon, puis chauffé pour obliger la double hélice à se séparer en brins individuels. La température est ensuite abaissée pour que l'ADN soit délimité en fragments de différentes tailles par des amorces d'ADN (courte séquence d'ADN à simple brin capable de s'hybrider avec l'ADN à séquencer). En augmentant légèrement la température et en ajoutant des nucléotides libres ainsi que des enzymes intervenant dans la réplication de l'ADN (ADN polymérase), les brins individuels vont pouvoir reformer leur brin complémentaire/inverse un nucléotide à la fois. Afin d'observer quel nucléotide a été sélectionné à chaque position et pour chaque fragment, les nucléotides utilisés ont été

préalablement modifiés pour identifier correctement la base selon leur type (A, T, C ou G). Pour pouvoir observer ce nouvel ajout de nucléotide, les fragments d'ADN sont séparés grâce à une électrophorèse en gel puis mis sous rayonnement UV (dans sa version la plus récente). Chaque piste correspond à un type de nucléotide (Sanger 1975). Tout cela se faisant "manuellement", il est alors compréhensible que le séquençage Sanger soit considéré comme fastidieux et coûteux. Aujourd'hui la méthode a été modifiée pour utiliser des fluorochromes capables d'émettre une lumière colorée spécifique pour chaque nucléotide utilisé par l'ADN polymérase (Smith et al 1985). L'analyse des séquences est effectuée par un laser relié à un ordinateur qui peut détecter la couleur renvoyée par le réactif dans son tube capillaire. Les données obtenues présentent une grande qualité avec seulement une base fautive toutes les 100.000 identifiées (taux d'erreur de 0,001%, sans compter les erreurs induites par la PCR en amont) et un temps d'analyse restant en dessous des 3 heures pour 1 à 5 nL d'ADN (Tableau 0-2).

Mais même automatisée, cette méthode possède beaucoup d'inconvénients. Après son prix, le deuxième point négatif est le volume de données produites qui est faible : quelques centaines de réactions seulement sont analysées en parallèle (avec un maximum de 384 réactions pour les meilleurs modèles de séquenceurs). Afin d'augmenter drastiquement ce volume, de nouvelles techniques de séquençage NGS ont fait leur apparition : SOLiD (Dekker 2007), Solexa (Service 2006) ou encore le pyroséquençage de 454 (Ronaghi et al 1998). L'ADN est fragmenté aléatoirement en séquences courtes mais de même taille. Ces millions de fragments sont ensuite parallèlement amplifiés et séquencés, les données obtenues sont nommées des lectures de séquences d'ADN (abrégées en lectures/*reads*). La technologie la plus utilisée à ce jour est celle de Illumina (Solexa avant 2017) avec plus de 90% de part du marché des machines de séquençage. Elles utilisent toutes une amplification clonale (PCR) puis un séquençage par synthèse (SBS) (Figure 0-10A). Une fois les fragments d'ADN simples brins obtenus par digestion de l'ADN, chacun d'entre eux est ensuite accroché par une seule de ses extrémités dans une cellule de flux (*flow cell*) puis amplifié. Plusieurs copies du fragment d'ADN se retrouvent alors collés sur cette plaque de verre, de la même manière que

le fragment initial, et séquencés en même temps par la technique vue précédemment. Des ADN polymérase reconstituent le complément du brin d'ADN collé et libèrent un groupe fluorescent à chaque nucléotide incorporé. Un laser couplé d'un ordinateur enregistrent ainsi les couleurs de chaque cellule de flux et reconstituent les lectures observées (Slatko et al 2018). Avec cette méthode, on obtient des paires de lectures (*paired reads* ou *pe*): la séquence du brin d'ADN et de son complément. Illumina permet aujourd'hui d'obtenir massivement des données ADN (milliards de lectures) tout en gardant un taux d'erreur faible (sans compter les erreurs induites par la PCR) (Tableau 0-2). Un score QPHRED est associé à chaque base détectée, et aide à évaluer la précision de l'identification de la base (score de qualité). Mais surtout, le coût d'un tel séquençage est très faible par rapport à celui de Sanger, bien qu'il nécessite plus de temps pour la même quantité d'ADN.

Les dernières technologies sorties tentent encore d'améliorer les données produites en proposant des lectures beaucoup plus longues que celles pouvant être obtenues par Sanger ou Illumina (Tableau 0-2). Ces séquençages de troisième génération (TGS, *long-reads*) regroupent deux méthodes distinctes : celle de Pacific Biosciences (PacBio ou PB) et celle de Oxford Nanopore Technologies (ONT). Dans le cas de la première, on reprend le principe des nucléotides fluorescents (Figure 0-10B). Cependant aucune amplification n'est nécessaire, les brins d'ADN doubles sont découpés en fragments de grande taille avant d'être circularisés et séquencés (Rhoads et Au 2015). Pour la deuxième méthode, il s'agit d'une technologie entièrement différente car non basée sur de l'ADN polymérase (Figure 0-10B). Un brin d'ADN simple est sélectionné pour traverser un nanopore un nucléotide à la fois (Wang et al 2021). La résistance détectée lors du passage variant selon le type de nucléotide, il sera alors possible de déterminer la séquence. Chacune de ces deux méthodes est alors capable de proposer un grand nombre de lectures en quelques heures. Mais l'obtention de lectures longues (*long-reads*) a un prix, les technologies TGS présentent un taux d'erreur modéré autour allant de 0,1 à 15% selon celle sélectionnée, et un coût légèrement plus élevé que Illumina. Mais ces inconvénients tendent à disparaître ou diminuer. Actuellement, ONT est descendu en dessous de la barre des 1% d'erreurs (méthode Q20/Q30+).

Chaque technologie a apporté son lot d'avantages et d'inconvénients au fil des années. A notre époque, il n'est pas rare de voir une reconstitution de génome à partir de données provenant de plusieurs de ces technologies, surtout lorsqu'il s'agit d'un assemblage *de novo* (reconstruction d'un génome sans connaissances préalables, sans pouvoir utiliser un génome de référence). Utiliser des lectures longues pour séquencer un génome dans sa globalité, puis des séquences courtes pour pouvoir corriger le génome assemblé, met à contribution les points forts de chacun : la couverture des lectures longues (longueur du génome séquencé) et la précision des lectures courtes.

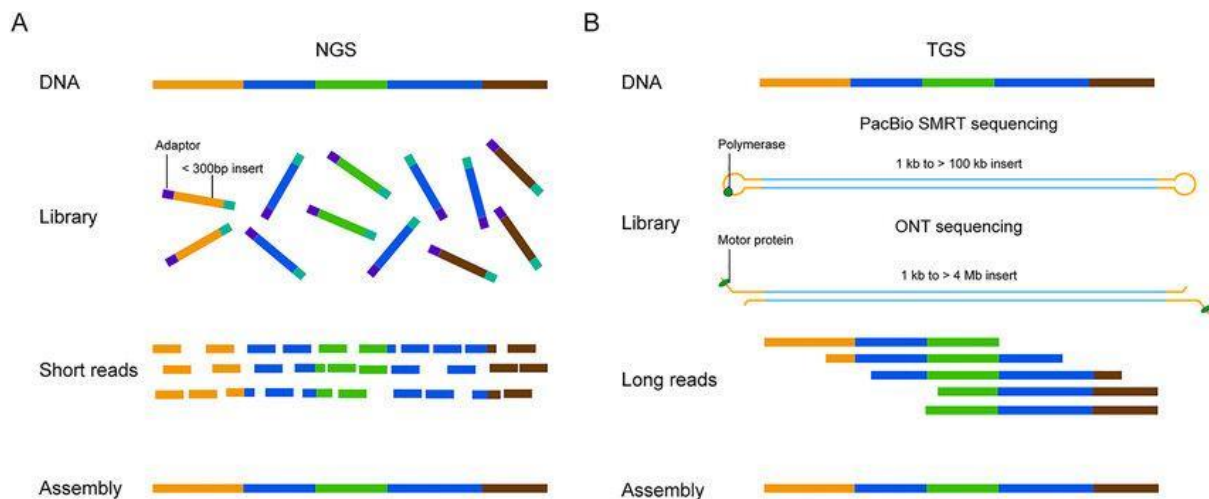


Figure 0-10 : Représentation schématique des méthodes de séquençage NGS (A) et TGS (B), Figure extraite de Chen et He (2021)

Les méthodes NGS (A) correspondent à celles utilisant des lectures courtes, et TGS (B) à celles usant de lectures longues.

“(A) Dans la technologie NGS d'Illumina, l'ADN est fragmenté en fragments de taille gérable et ces fragments sont ligaturés à des adaptateurs. Après la préparation de la bibliothèque, les molécules d'ADN individuelles sont séquencées pour les lectures courtes. Après un séquençage, les données brutes sont alignées sur un génome de référence. (B) Dans le séquençage SMRT de PacBio, le fragment d'ADN est ligaturé à des adaptateurs en épingle à cheveux pour former une molécule topologiquement circulaire, connue sous le nom de SMRTbell. Celle-ci est chargée sur une cellule SMRT et liée par une ADN polymérase pour le séquençage. Dans le séquençage ONT, l'ADN est marqué par des adaptateurs de séquençage préchargés d'une protéine motrice à l'une ou aux deux extrémités. L'ADN est combiné avec des protéines d'attache et chargé sur la cellule de flux pour le séquençage. À la suite d'un séquençage, les séquences brutes sont alignées sur un génome de référence. NGS, séquençage de nouvelle génération ; TGS, séquençage de troisième génération.”, description traduite de Chen et He (2021).

Nom	Génération	Année de sortie	Taux d'erreur (%)	Longueur des lectures (bases)	Prix moyen pour un millions de bases (\$USA)	Nombre de lectures par analyse	Temps par analyse
Sanger	1	1975	0.001%	300 à 850b	2,400\$	N/A	20min à 3h
Illumina NextSeq	2	2017 (2006)	0.1%	en paires de 50, 150 ou 250b	0.05 à 0.15\$	<= 3.000.000.000	1 à 4 jours
Pacific Biosciences Single-molecule real-time (SMRT)	3	2010	10 à 15% (1% en HiFi)	10.000 à 18.000b	10\$	<= 75.000	1 à 2 jours
Oxford Nanopore Technologies (ONT)	3	2014	1 à 5% (0,5 Q20+)	20.000 à 2.000.000b	15\$	<= 80.000	17h à 3 jours

Tableau 0-2 : Tableau récapitulatif des informations sur les principales techniques de séquençage utilisées en Septembre 2023

Toutes les informations proviennent des sources suivantes : Deschamps et al (2012), Donkor (2013), Villain (2018), www.pacb.com, www.nanoporetech.com, www.illumina.com, Vilella v1.10.28 (2023).

3. Les méthodes d'assemblage et de métassemblage

Toutes les technologies de séquençage vues précédemment permettent d'obtenir des fragments plus ou moins grands d'ADN. Il est donc nécessaire d'assembler ces morceaux de séquence pour pouvoir reconstituer le génome de l'espèce considérée. On appelle cette étape l'assemblage du génome. De nos jours, cette reconstitution du génome se fait par informatique et diffère selon les données de séquençage.

Dans le cas de lectures courtes, le volume de données produites étant massif, il faudrait des années pour tester toutes les combinaisons possibles en comparant les fragments deux à deux sans méthode de calcul, et ce même avec l'ordinateur le plus puissant au monde. Deux grandes approches permettent de faciliter les calculs, les parcours Hamiltonien et

Eulérien (Compeau et Pevzner 2015), en découpant les lectures en k-mers, c'est-à-dire un ensemble de mots de k lettres que l'on peut lire dans une séquence d'ADN. Par exemple, la séquence ATCGTT est constituée de trois 4-mers : ATCG, TCGT et CGTT (Figure 0-11a). Seul le parcours Eulérien sera détaillé ci-dessous, puisqu'il est le seul pour lequel il existe un algorithme informatique assez rapide pour le rechercher dans le graphe des données de séquençage. Ainsi un graphe orienté est construit grâce aux k-mers, autrement dit un ensemble de nœuds reliés par des flèches, où chaque k-mer représente une flèche. Les suffixes et préfixes d'un k-mer, et de taille k-1, sont contenus dans les nœuds en amont et en aval de la flèche, respectivement (Figure 0-11b, graphe de Bruijn). Les nœuds identiques doivent être fusionnés pour éviter les complications pour la suite. Le graphe étant constitué de tous les k-mers contenus dans les lectures séquencées, il "suffit" alors de tracer un chemin passant une unique fois par toutes les flèches du graphe pour retrouver le génome d'origine (parcours Eulérien). Beaucoup de logiciels d'assemblage de lectures courtes ont été créés à travers les années, mais pour ne citer que les plus connus : SPAdes (Bankevich et al 2012), Velvet (Zerbino et al 2008) et ABySS (Simpson et al 2009). Tous les trois utilisent le parcours Eulérien et le graphe de Bruijn.

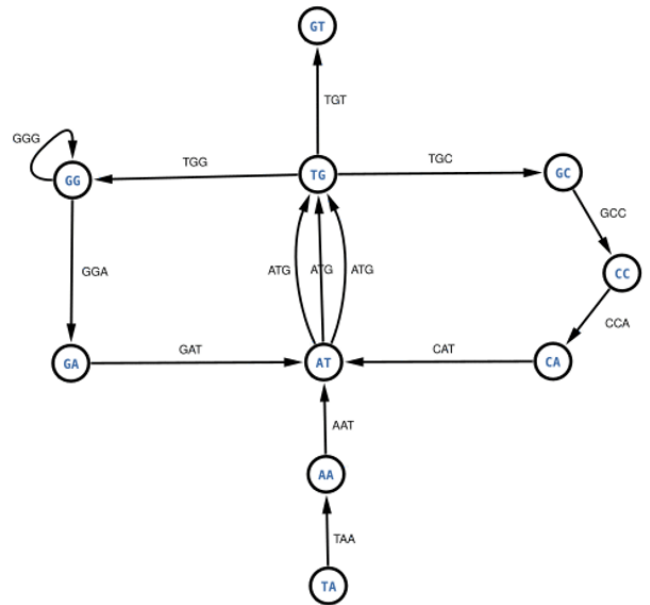
(A)

TAATGCCATGGGATGT
 TAA TGCCATGGGATGT
 AAT GCCATGGGATGT
 TAATGCCATGGGATGT
 TAATGCCATGGGATGT
 TAATGCCATGGGATGT
 ...
 TAATGCCATGGGATGT

Légende :

- | | | | |
|------|------------------|------|-------------------------------|
| ATCG | Séquence ADN | ATCG | 3-mer formé par le chemin |
| ATCG | 3-mers extrait | ATCG | Préfixe et Suffixe des 3-mers |
| ATCG | Séquence ignorée | → | Chemin dans le graphe |

(B)



TAATGCCATGGGATGT

Figure 0-11 : Représentation schématique des k-mers extraits d’une séquence ADN (A) et du graphe de Bruijn correspondant (B)

Une valeur de k=3 est prise ici pour exemple, il s’agit donc de 3-mers. Pour récupérer ces derniers, une fenêtre de 3 lettres glisse par pas de 1 lettre.

L’image (B) est extraite de <https://dridk.me/assemblage.html>. Pour retrouver la séquence ADN donnée, il est nécessaire de parcourir ce graphe de Bruijn de manière à former un parcours Eulérien, donc de passer par toutes les flèches du graphe.

Cette méthode devient cependant obsolète avec l’apparition des séquençages de troisième génération. Les lectures longues entières sont alors préférées à l’utilisation des k-mers qui présentent de nombreux désavantages. Leur inconvénient principal concerne les régions répétées du génome: si la taille du k-mer n’est pas plus importante que celle des répétitions, alors il sera impossible de retrouver un chemin dans le graphe. Le génome se retrouve alors divisé en autant de fragments que de répétitions non résolues, morceaux appelés des contigs. Ce problème est peu rencontré avec les lectures longues, puisqu’elles peuvent mesurer plusieurs milliers de bases, et permettent d’obtenir des morceaux contenant les répétitions en entier et entourées de séquences non répétées. De plus, les séquences

d'ADN ne sont jamais passées par une étude d'amplification, ce qui leur confère encore un avantage par rapport aux lectures courtes, aucun biais de profondeur n'ayant été introduit dans les données (calcul d'abondance possible). Enfin, des techniques existent pour diminuer le taux d'erreur des données de troisième génération. Les lectures erronées peuvent être corrigées de deux façons : soit en alignant les lectures longues entre elles avant l'assemblage, soit en alignant des lectures courtes ou longues sur les contigs ou génome obtenus à partir des lectures longues (*polishing step* de l'assemblage). Dans les deux cas, on considérera que le nucléotide réel sera celui dont la profondeur (nombre de fois où une position est séquencée, mesurée en X) est la plus grande. Actuellement, le meilleur logiciel d'assemblage de lectures longues semble être Flye (Kolmogorov et al 2019), d'après plusieurs comparatifs avec les logiciels les plus utilisés (tel que Canu, Koren et al 2017), et réalisés sur des génomes différents (Wick et Holt 2019, Zhang et al 2022). Il utilise un algorithme basé sur un graphe de répétitions, lui-même construit à partir de correspondances de séquences approximatives. Ainsi, le logiciel ne cherche pas une séquence exacte comme précédemment, mais une séquence approximative, ce qui permet de prendre en compte le taux d'erreur des lectures longues. Les flèches du graphe représentent une séquence d'ADN, pendant que les nœuds modélisent les jonctions. Chaque flèche est classée en flèche unique ou en flèche répétitive. Le logiciel calcule ensuite le chemin ne passant qu'une unique fois par chaque flèche du graphe (Figure 0-12).

Il est à noter qu'une fois le génome obtenu par l'une de ces méthodes, il est tout de même nécessaire de vérifier que ce dernier soit bien complet et correctement assemblé. La méthode la plus courante pour cela est d'utiliser le logiciel BUSCO (Manni et al 2021). A partir de la base de données orthoDB (Kuznetsov et al 2022), BUSCO recherche dans le génome nouvellement assemblé des séquences consensus de gènes orthologues spécifiques à de grands groupes taxonomiques. Constitués d'un large éventail d'espèces pour représenter au mieux leur diversité, ces taxons sont utilisés pour construire des séquences consensus que toute espèce leur appartenant doit posséder. Ainsi un score BUSCO est calculé selon les gènes identifiés par le logiciel et permet de s'assurer de la complétude d'un assemblage.

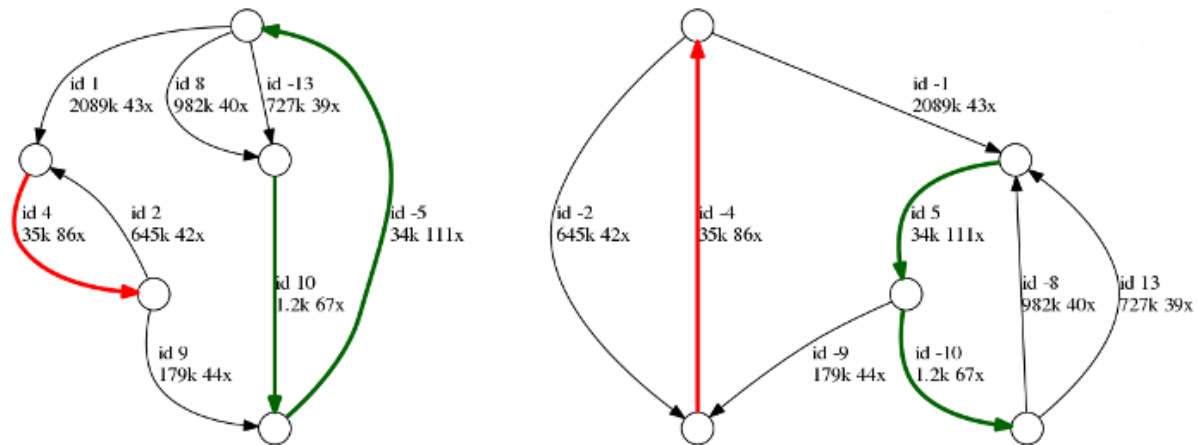


Figure 0-12 : Exemple d'un graphe de répétition d'un assemblage bactérien réalisé par le logiciel Flye, Figure extraite du Github de Flye

Les arêtes sont associées à leur identifiant unique (id), leur longueur en milliers de bases (k) et leur profondeur de séquençage (x). Celles répétitives sont en couleur tandis que les arêtes uniques sont en noir. Chaque arête est schématisée deux fois pour représenter le brin direct (+) et son complément (-), ce qui signifie que le génome entier est visible ici en deux copies.

L'image est disponible sur : <https://github.com/fenderglass/Flye/blob/flye/docs/USAGE.md> .

Mais toutes ces méthodes ont été décrites dans le cas d'un assemblage classique, c'est-à-dire avec un seul organisme séquencé. Il est tout à fait possible que les échantillons séquencés soient constitués de plusieurs êtres vivants. On parle alors de séquençage métagénomique et de métassemblage. Les lectures longues récupérées proviendront de plusieurs organismes différents (eucaryotes et/ou procaryotes), sans marquage quant à leur origine. Ces séquences mélangées les unes aux autres complexifient grandement l'étape d'assemblage, il ne sera alors plus question de reconstituer le génome d'un seul être vivant mais celui d'un nombre d'organismes, souvent inconnu. De nouveaux logiciels sont utilisés pour prendre cette particularité en compte, tout comme Flye qui possède une version pour la métagénomique : MetaFlye (Kolmogorov et al 2020). Des logiciels existent également pour les lectures courtes, mais dans le cas de métagénome il sera préféré l'utilisation des lectures longues corrigées par lectures courtes pour obtenir des génomes de meilleure qualité (Chen, L. et al 2022), même si cette tendance semble s'estomper actuellement avec la diminution du taux d'erreur des séquençages ONT (Latorre-Pérez et al 2020). Mais bien que plus complexe à réaliser, le séquençage métagénomique et le métassemblage présentent de nombreux

avantages par rapport aux méthodes classiques. Son plus grand avantage est le séquençage de presque tous les organismes présents au sein de l'échantillon étudié, quelle que soit leur concentration dans ce dernier. Sans sélection préalable ni amplification de l'ADN, les organismes les plus abondants dans l'échantillon auront plus de chance d'être séquencés et auront une grande profondeur de séquençage. On peut donc récupérer toutes les espèces présentes, mais également estimer leur abondance dans les échantillons (Yorki et al 2023). Cette méthode semble être très intéressante dans le cas des diatomées non cultivées de manière axénique. Le séquençage de leur culture contiendra de l'ADN d'organismes eucaryotes (la diatomée) et procaryotes (bactéries) qu'il faudra assembler en parallèle pour pouvoir en extraire le génome de la microalgue.

4. L'élimination des génomes "contaminants", une des limites de la métagénomique

Le cas le plus simple pour identifier un génome spécifique dans un métassemblage est d'utiliser un génome de référence. Si des espèces proches de celle d'intérêt ont déjà été assemblées auparavant, il est possible de récupérer leurs séquences et d'aligner l'entièreté du métassemblage sur ces dernières. Cet alignement permettra d'identifier les contigs homologues aux données de référence en filtrant les résultats pour n'obtenir que des alignements avec un pourcentage d'identité de séquence (%ID, pourcentage de résidus similaires dans un alignement) et un pourcentage de couverture au-dessus d'un certain seuil (souvent 80%). Des bases de données existent et regroupent un grand nombre de génomes déjà annotés taxonomiquement, comme la base de données du NCBI nr/nt (*RefSeq non-redundant nucleotide*) et celle de UniProt swissprot (séquences de protéines). A noter que cette méthode peut aussi être utilisée à l'inverse. Si les bases de données ne contiennent aucun génome pouvant être utilisé comme référence pour l'organisme d'intérêt, alors il est

possible d'essayer d'identifier les autres contigs du métassemblage pour réduire le nombre de séquences appartenant potentiellement à ce dernier.

Mais dans certaines conditions, l'identification de génome d'intérêt dans un métassemblage peut représenter un défi de taille. Ce problème est d'ailleurs souvent cité comme une des grandes limitations de la métagénomique pour laquelle il est difficile de trouver des solutions (Bharti et Grimm 2021). Les méthodes décrites précédemment ou ci-dessous peuvent paraître simples et faciles à mettre en place, cependant leur application en conditions réelles s'avèrent souvent ardues et complexes. Les exemples sur le sujet sont nombreux, mais peu d'entre eux sont publiés comme il s'agit de résultats négatifs, et sont donc retrouvés dans les forums de bioinformaticiens tels que BioStars. Cependant, un autre indicateur de la difficulté de cette étape peut être le nombre de logiciels sortant chaque année pour résoudre cette limitation (Constantinides et al 2023, Jiang et al 2023, Lataretu et al 2023, Mangalea et al 2023). Des soucis de décontamination de métagénomiques sont retrouvés dans de nombreux cas différents : une très grande diversité des espèces présentes dans l'échantillon de départ, si ces dernières partagent une grande similarité entre leurs séquences, ou encore si les contigs reconstitués sont de trop petite taille pour être identifiés correctement, etc. Il est souvent nécessaire d'utiliser des méthodes spécifiques pour chaque cas de figure, mais sans assurance que ces dernières fonctionnent comme souhaité.

Par exemple, la méthode précédente ne fonctionne que si l'on considère que les organismes présents dans les échantillons ont déjà été assemblés auparavant. Dans le cas des diatomées et de *H. ostrearia*, seuls trois génomes ont déjà été reconstruits et aucun d'entre eux ne semble pouvoir être utilisé comme génome de référence (Tableau 0-1, *Thalassiosira pseudonana* Armbrust et al 2004, *Phaeodactylum tricornutum* Bowler et al 2008, *Pseudo-nitzschia multiseriis* JGI 2011). De plus, les seuls microorganismes marins connus à ce jour sont ceux jouant un rôle dans les processus biogéochimiques (Sogin et al 2006). Il y a de grandes probabilités que les bases de données ne contiennent pas ou très peu d'informations sur les bactéries peu abondantes ou difficilement cultivables. Il faudra donc

utiliser d'autres méthodes pour annoter taxonomiquement chacun des contigs d'un métassemblage. Il est possible, par exemple, d'utiliser des caractéristiques partagées par un, plusieurs ou tous les eucaryotes ou procaryotes pour cela.

Si les séquences assemblées sont assez longues, il est possible d'utiliser le %GC moyen, la composition en éléments transposables (ETs) ou encore la recherche de gènes conservés, par exemple. Les génomes bactériens possèdent un %GC moyen variant entre 25% et 75% (Sueoka 1962), voire même en dessous de 20% (Nakabachi et al 2006), contrairement aux diatomées qui semblent graviter autour des 40-50%. Il sera également possible de rechercher des classes d'ETs uniquement présents chez les eucaryotes, comme les rétrotransposons (Classe I). Tous les génomes eucaryotes connus à ce jour possèdent une quantité d'ETs variable selon l'espèce, et il n'existe que peu d'exceptions à cette règle (Wells et Feschotte 2020). Enfin, chaque espèce eucaryote ou procaryote possède des gènes qui sont conservés au fil des générations (gènes marqueurs, séquences consensus, etc). C'est pourquoi le gène ARNr 16S et 18S sont très utilisés par la communauté scientifique. Ces gènes, respectivement procaryotes et eucaryotes, possèdent des régions qui sont fortement conservées au sein de leur domaine et des régions hypervariables qui varient selon l'espèce/la famille (Chakravorty et al 2008, Hadziavdic et al 2014, Figure 0-13). De plus, suite à leur grande utilisation en phylogénie, il existe des bases de données importantes regroupant ces informations, comme SILVA (Quast et al 2012).

Mais les contigs peuvent être mal assemblés et être ainsi trop courts pour rechercher des caractéristiques spécifiques à un taxon. Ce cas de figure peut se présenter lorsque les échantillons contiennent une grande diversité d'organismes dont certains sont très peu abondants. Les informations sur ces derniers seront "noyées" dans la masse de données et les logiciels d'assemblage ne seront pas en capacité de reconstruire leur séquence correctement. Leur identification taxonomique devient complexe voire impossible. D'autant plus lorsque le métagénome est construit sans *a priori*, et que le nombre de génomes le composant n'est pas connu. Il est difficile de savoir si les morceaux restants, après l'élimination

des séquences des espèces abondantes *via* les méthodes précédentes, n'appartiennent qu'à une seule espèce ou plusieurs.

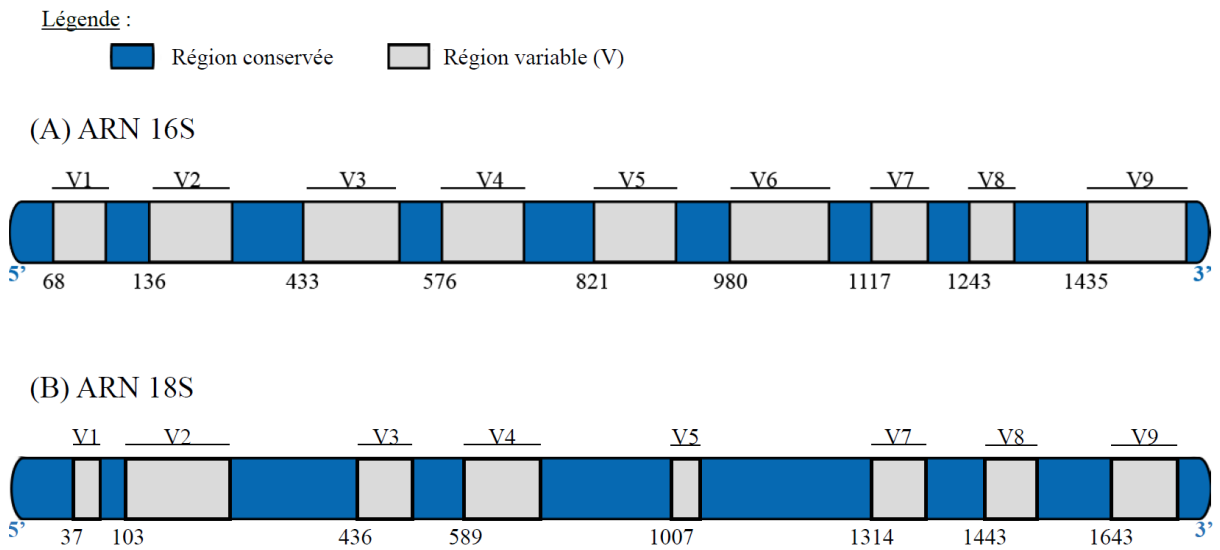


Figure 0-13 : Représentation schématique des ARNs ribosomiques 16S (A) et 18S (B)

La région V6 de l'ARN 18S n'est pas représentée ici puisqu'elle est peu variable en comparaison des autres régions, et est donc souvent ignorée par la communauté scientifique (Neefs et al 1993).

Partie 3 : L'antibiorésistance, un problème mondial

"Les antibiotiques c'est pas automatique." - Campagne de la Caisse nationale d'assurance maladie (2002)

1. Depuis quand parle-t-on d'antibiorésistance ?

Au début du XXe siècle, une substance produite par des bactéries vivant dans le sol est découverte par le biologiste français Dubos (Avery et Dubos 1930). Il parvient à l'isoler en 1937 et la nomme la gramicidine pour ses capacités à inhiber la croissance de différentes bactéries à gram positif (Dubos 1937). Il a ainsi été à l'origine du premier antibiotique naturel et testé sur l'être humain pendant la Seconde Guerre mondiale : la tyrothricine. Cette découverte, bien que vite abandonnée à cause de sa grande toxicité (Rammelkamp et Weinstein 1942), a inspiré les travaux de 1940 à l'origine de l'antibiotique souvent cité comme le premier mondial : la pénicilline. En effet, en parallèle des recherches de Dubos, Fleming a découvert une substance synthétisée par *Penicillium notatum*, une moisissure, qui était capable d'inhiber la croissance de bactéries. C'est en concentrant cet antibactérien naturel qu'il a inventé en 1928 une première forme instable de ce dernier : la pénicilline (Fleming et al 1928, Figure 0-14). Fleming n'a pas été à l'origine de la démocratisation et de l'utilisation de cet antibiotique dans le traitement de pathogènes humains. Ce n'est qu'à partir de 1940 que les premiers essais de la pénicilline sur des blessés britanniques ont été réalisés grâce aux travaux de Florey et Chain (Ligon 2004, Figure 0-14). Depuis 1946, les antibiotiques peuvent se trouver en vente libre dans les pharmacies des pays développés et représentent un marché florissant pour les laboratoires pharmaceutiques (Aminov 2010). Une centaine de molécules différentes et antibactériennes apparaissent et sont utilisées librement par la population mondiale. Cette utilisation, aujourd'hui qualifiée d'abusives, a engendré la première mise en évidence de bactéries pathogènes capables de résister aux antibiotiques usuels au Japon

(Davies 1995, Figure 0-14). Il s'agit de la première fois que la notion d'antibiorésistance apparaît dans des recherches.

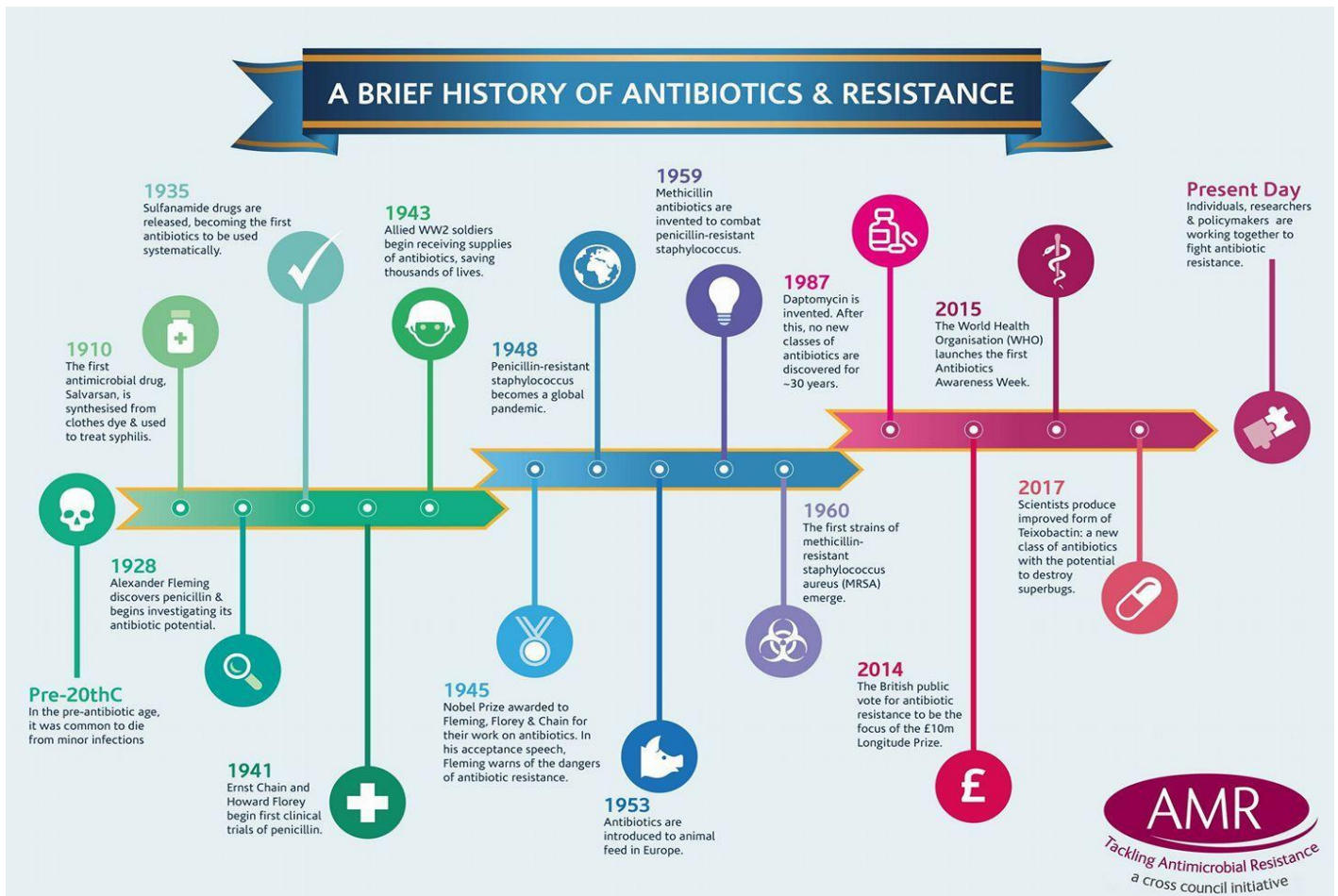


Figure 0-14 : Frise chronologique des évènements marquants dans l'histoire du phénomène d'antibiorésistance depuis la création des antibiotiques, Figure extraite d'une communication digitale de 2017 de UK Research and Innovation

Image disponible sur : https://x.com/UKRI_News/status/930003013489627136?s=20.

2. Les enjeux de l'antibiorésistance

L'antibiorésistance se définit comme l'acquisition d'une résistance partielle ou totale à un antibiotique par une bactérie qui est habituellement sensible à ce dernier. Cette capacité rend alors l'usage de cet anti-infectieux partiellement ou totalement inefficace dans le cadre

médical (Watkins et Bonomo 2016), vétérinaire (Teuber 2001) ou le traitement de cultures en laboratoire (par exemple d'algues avec Jones et al 1973).

C'est ainsi qu'en 1948, une première pandémie de bactéries résistantes aux antibiotiques dérivés de la pénicilline a été constatée (Figure 0-14, Barber et Rozwadowska-Dowzenko 1948). Cette antibiorésistance ayant été acquise par la mise en contact répétée de *Staphylococcus aureus* à ce type de molécules antibactériennes. Pour lutter contre cette pandémie, une nouvelle molécule semi-synthétique a alors été inventée, la méthicilline, en 1959 (Enright et al 2002). Mais cette dernière n'a fait que renforcer ses capacités à résister aux antibiotiques en lui donnant la possibilité d'en acquérir face à celle-ci (Jevons 1961). En effet, l'un des principaux facteurs qui favorisent l'émergence de bactéries résistantes est l'utilisation répétée d'antibiotiques (pression de sélection), et ce peu importe le cadre dans lequel cet usage est fait (Teuber 2001, Aminov 2009). Mais ce problème n'a été considéré comme tel que lorsqu'il a été découvert que ce phénomène d'acquisition d'antibiorésistance ne s'arrêtait pas seulement à *Staphylococcus aureus*. Des cultures de *Escherichia coli* se sont montrées également insensibles à certains antibiotiques (Echeverria et al 1978, Ghenghesh et al 2013), de même pour le genre *Shigella* (Davies et al 1970). Le phénomène semble être généralisé à travers le monde, et toucher un large panel d'espèces bactériennes.

Des travaux ont alors été menés pour mieux comprendre cette émergence de résistances et lutter contre ce problème grandissant. En 1963, Watanabe démontre que l'antibiorésistance est la conséquence de gènes (ARGs) présents dans les plasmides bactériens, alors appelés "épisomes" (Watanabe 1963). Dans cet article, il explique comment des bactéries considérées inoffensives, telle que *Escherichia coli* (bactérie intestinale très souvent mise en contact avec différents antibiotiques), peuvent facilement transmettre leurs résistances à d'autres espèces bactériennes pathogènes. En effet, un plasmide bactérien peut être échangé entre deux bactéries, et ce même si ces dernières n'appartiennent pas au même genre ou à la même famille (conjugaison bactérienne, Frost 1992). Il suffit alors que ce dernier contienne de nombreuses résistances pour qu'une souche receveuse du plasmide et sensible à de nombreux antibiotiques ne le soit plus. Ce mécanisme évolutif basé sur l'échange ne

semble cependant pas être nouveau, une analyse sur plusieurs groupes de ARGs suggère que ces derniers ont connu une longue histoire de sélection et de diversification, et ce en amont de “l’âge d’or” des antibiotiques (Aminov et Mackie 2007). Cette émergence de résistances aux antibiotiques a simplement été intensifiée par la découverte et l’invention d’antibiotiques toujours plus diversifiés et abondants.

Le début des années 2000 est alors marqué par une prise de conscience collective de l’importance de réduire l’utilisation d’antibiotiques au niveau mondial. En effet, bien que de nouveaux produits soient synthétisés par les entreprises pharmaceutiques, ces derniers deviennent de plus en plus limités. Les nouvelles molécules se raréfient en même temps que les ARGs s’accumulent dans les génomes bactériens; l’arsenal thérapeutique disponible pour lutter contre les bactéries pathogènes diminue alors dangereusement. Il aura fallu, de plus, attendre 2000 avant la fabrication du premier antibiotique de nouvelle génération (Linezolide, Clemett et Markham 2000), soit 20 ans après les précédents. Ainsi, pour éviter l’apparition de bactéries pathogènes multirésistantes, l’accès aux antibiotiques est alors limité, et de nombreuses campagnes publicitaires voient le jour pour sensibiliser et informer la population mondiale sur cette diminution de l’effet des antibiotiques, et la nécessité de les utiliser correctement (Figure 0-14, Cohen 2004, Anjuli 2019, Harbarth et al 2015). Par exemple, dans le cadre médical, il était fréquent que des antibiotiques soient donnés à des patients atteints d’infections virales (les antibiotiques n’ont pas d’effet sur les virus, McNulty et al 2007), ou durant une durée non optimale (l’hôte possédait encore des bactéries qui ont survécu aux molécules et qui ont pu muter en conséquence) (Ding et He 2010). En parallèle, en 2013, une liste d’antibiotiques de “dernier recours” ou “critiques” a été mise en place, notamment en France par la Haute Autorité de Santé (HAS). Cette dernière encadre l’utilisation de molécules pour lesquelles il existe encore peu de résistances chez les bactéries, afin de conserver leur effet sur le long terme. Les antibiotiques de “dernier recours” ne sont alors utilisables qu’en cas extrême, dont les conditions sont disponibles dans les rapports de la HAS (HAS 2014). En 2014, l’Organisation Mondiale de la Santé (OMS) sort également un rapport sur l’état critique

de l'antibiorésistance mondial (World Health Organization 2014), et lance sa première campagne autour de cette prise de conscience (Harbarth et al 2015).

Aujourd'hui encore, de nouvelles résistances sont retrouvées dans les bactéries et représentent une menace mondiale pour la santé des êtres humains (World Health Organization 2007), des animaux d'élevage (Mo et al 2017, Tiedje et al 2023), mais également de l'agriculture (Tiedje et al 2023, Wu et al 2023). Il s'agit d'un sujet fortement traité chaque année et dans de nombreux domaines, et pour lequel la seule solution viable actuellement proposée est la réduction de l'utilisation des antibiotiques.

3. La recherche des marqueurs d'une potentielle résistance aux antibiotiques

Devant ce problème de plus en plus important mondialement, il est alors nécessaire de pouvoir détecter les gènes responsables de ces résistances en étudiant directement les génomes bactériens. Cette étude génomique permet d'estimer la menace que pourrait représenter une bactérie, qu'elle soit pathogène ou non, plus efficacement et rapidement que des méthodes plus "manuelles" (antibiogrammes, Truong et al 2021) (Palmer et Kishony 2013). Par exemple, en utilisant des techniques basées sur le séquençage d'ADN, il est ainsi possible de rechercher un large panel de ARGs différents dans un grand nombre de génomes bactériens (Gupta et al 2014, McArthur et Wright 2015), voire même de détecter des résistances de type sauvage, dont les gènes ne sont pas exprimés par les cellules bactériennes et qui pourraient ainsi ne pas être observées par antibiogrammes (Enne et Bennett 2010).

L'une des méthodes les plus simples s'appuie sur un séquençage de gènes amplifiés par PCR. Grâce aux précédentes recherches, un nombre réduit de gènes ont été identifiés comme responsables de la résistance à plusieurs antibiotiques et ce chez une grande diversité de bactéries. Des amorces ont alors pu être proposées pour vérifier la présence de chacun de

ces gènes (Lina et al 1999, Cockerill 1999). C'est ainsi que 9 gènes ont, par exemple, été recherchés en simultanée chez une bactérie pathogène de l'être humain par une PCR multiplex (Strommenger et al 2003). Aujourd'hui, la société Illumina propose même un service de séquençage visant 478 gènes de résistances aux antimicrobiens (AMRs) différents, ce qui correspond à une insensibilité à 28 classes d'antibiotiques différentes (*AmpliSeq for Illumina Antimicrobial Resistance Research Panel*, référence 20020495 sur emea.illumina.com). Cependant, cette technique nécessite l'utilisation d'amorces déjà identifiées et vendues sur le marché. Ce besoin entraîne donc des biais, car les ARGs recherchés sont sélectionnés en amont des tests PCR.

D'autres méthodes plus complexes sont alors possibles pour pallier ce problème. Ainsi, il est possible d'utiliser directement le génome complet d'une ou plusieurs bactéries pour procéder à l'identification des gènes conférant une insensibilité ou une résistance à un ou plusieurs antibiotiques. Cette étude peut être effectuée de plusieurs façons, toutes reposant sur l'utilisation d'une base de données de ARGs. Le ou les génomes pourront soit être découpés en k-mers comparés à des k-mers signatures de résistances à certains antibiotiques (Jaillard et al 2020), soit être analysés pour retrouver des similarités entre les séquences protéiques ou nucléiques de leurs gènes et de ceux d'autres bactéries déjà associés à de l'antibiorésistance (McArthur et al 2013, Alcock et al 2023). Certains logiciels, comme RGI (McArthur et al 2013), utilisés avec la bonne base de données, telle que les différentes versions de CARD (Alcock et al 2023), permettent d'utiliser facilement ces deux approches et de produire de meilleurs résultats (Figure 0-15). Cette méthode basée sur le génome entier présente l'avantage de ne pas sélectionner en amont les ARGs, et ainsi de proposer une annotation "complète" de ces derniers (qui s'améliora en même temps que les dernières découvertes d'ARGs). Toutefois, elle nécessite tout de même d'utiliser une base de données adaptée au génome bactérien passé au crible. Également, suite à la complexité pour appliquer cette dernière, de solides connaissances en bioinformatique sont nécessaires contrairement à la première technique citée.

In fine, ces deux méthodes permettent d'obtenir la localisation dans le génome des ARGs, mais également d'avoir une idée de la facilité de transmission de ces derniers (gènes contenus dans le génome nucléaire ou plasmidique). Avec les années, ces analyses deviennent de plus en plus précises et semblent se détacher de la nécessité d'être également testées sur le niveau phénotypique (Maunsell et al 2021). En effet, par le passé, elles ont souvent été critiquées et qualifiées de "résultats non définitifs", car trop hypothétiques contrairement aux tests réalisés sur le phénotype (Cockerill 1999).

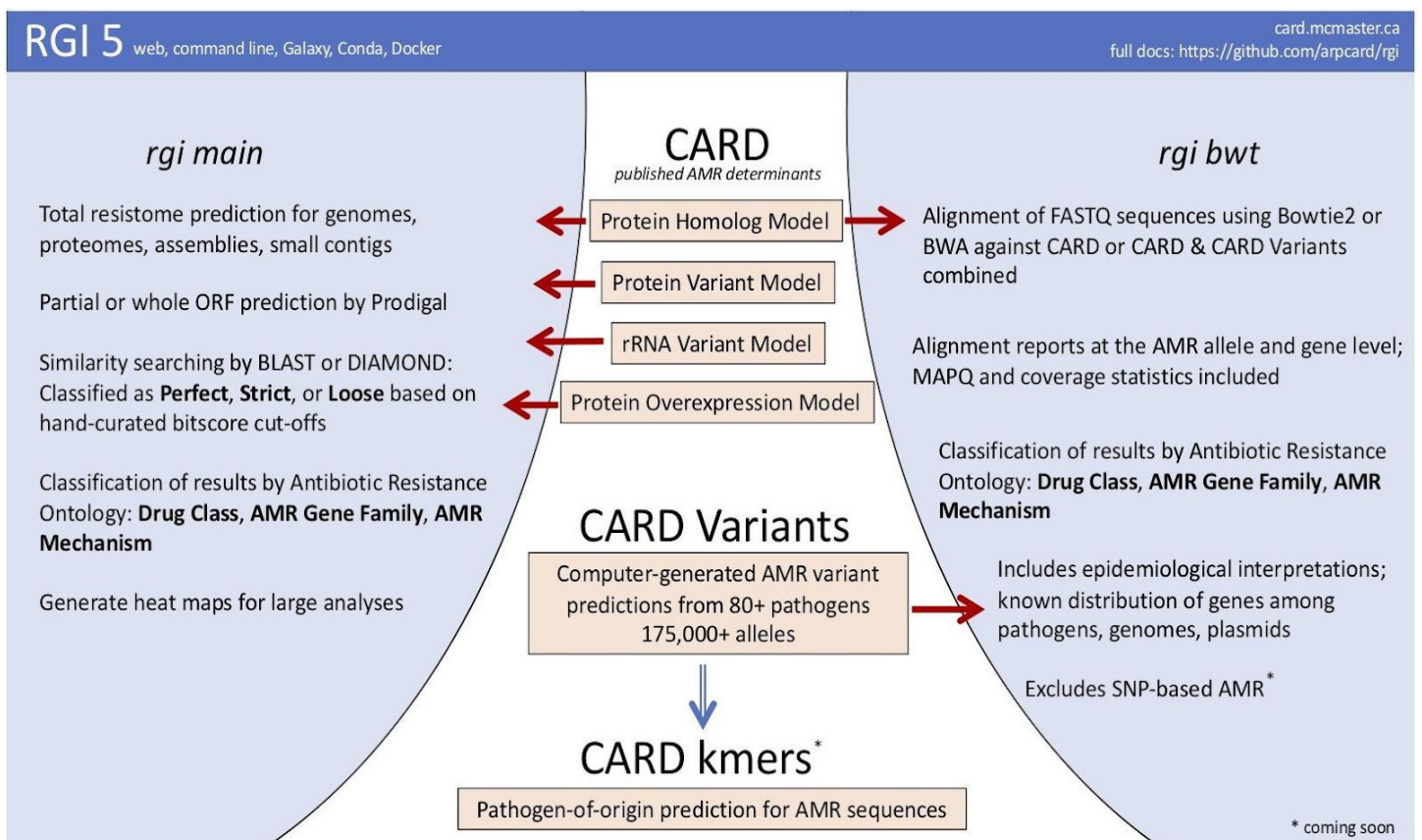


Figure 0-15 : Schéma explicatif du fonctionnement du logiciel RGI et des résultats obtenus avec l'utilisation de la base de données CARD, Figure extraite d'une communication orale de Maguire et al.

Disponible sur <https://finlaymagui.re/>.

Chapitre 1 : A la recherche du génome de *H. ostrearia*

1. Contexte des travaux de recherche

Au commencement de la présente thèse en octobre 2020, l'objectif principal était la reconstruction et l'étude de la composition en gènes du génome de *H. ostrearia*, avec un intérêt particulier pour les longs ARN non codants. Pour obtenir ce génome, un isolat récupéré le 16 janvier 2018 dans un bassin d'huîtres situé à Bouin en France (46.953444, -2.046139, Poulin et al. 2019) a été cultivé au laboratoire, puis déposé à la Nantes Culture Collection et est aujourd'hui conservé sous le nom de NCC 532 (contact : vona.meleder@univ-nantes.fr ; Nantes, France). Plusieurs séquençages ont été réalisés au laboratoire : deux séquençages en lectures longues, grâce à la technologie Nanopore (Gridlon, ONT) et Pacific Biosciences (CLR SEQUEL2, PacBio), ainsi que deux séquençage en lectures courtes grâce à Illumina MiSeq afin de corriger ces premiers (Tableau 1-1).

Type de Séquençage	Type des Données	Taille des Lectures	Date	Plateforme de Séquençage	Quantité de Lectures Générée	Taille Totale des Données
Nanopore Gridlon (ONT)	ADN	4 à 151 kb	Septembre 2020	Genotoul, Toulouse	2.9M	18,520 Gb
Illumina MiSeq TruSeqv3	ADN	2*150 b	Septembre 2020	Genotoul, Toulouse	17,47G x 2	5.241Gb
Illumina NovaSeq TruSeq	ARN poly-A	2 jeux de données de 2*150 b	Février 2021	Genotoul, Toulouse	188,45G x 2 par jeux de données	113.070Gb
PacBio CLR SEQUEL2 (PB)	ADN	10 à 50 kb	Septembre 2021	Genotoul, Toulouse	N/A	85,25Gb

Tableau 1-1 : Séquençages réalisés sur l'isolat NCC 532 de *Haslea ostrearia* au laboratoire BIOSSE entre 2020 et 2021 inclus

Pour appuyer ces travaux, les seules données disponibles en ligne sur l'ADN de *H. ostrearia* étaient constituées des séquences partielles appartenant au génome mitochondrial, gène *COI* pour "Cytochrome Oxidase Subunit 1", et au génome chloroplastique avec le gène *rbcL* "Ribulose bisphosphate carboxylase à large chaîne" (base de données du NCBI). Les seules informations complètes sur le génome des diatomées provenaient de trois espèces de diatomées non bleues dont les génomes publiés, respectivement en 2008 et en 2004, de *Phaeodactylum tricornutum* (Bowler et al 2008) et de *Thalassiosira pseudonana* (Armbrust et al 2004), ainsi que celui non publié de *Pseudo-nitzschia multiseriis* (JGI 2011). Malheureusement, le classement taxonomique des diatomées est souvent remis en question au fil des découvertes (Round et al 1990, Medlin and Kaczmarek 2004, Ruggiero et al 2015), il est donc difficile d'évaluer la distance évolutive entre ces dernières et *H. ostrearia*, surtout avec peu de données collectées sur le génome de cette dernière. Il était alors compliqué d'utiliser leur génome comme référence pour reconstruire celui qui nous intéressait, le risque d'y incorporer des erreurs et malfaçons d'assemblage étant trop important.

Il a été également rapporté que de petites colonies bactériennes ont été observées en laboratoire lors de l'étalage sur boîtes de Pétri des cultures de *H. ostrearia*. Il a alors été émise l'hypothèse que les données de séquençage obtenues contiendraient quelques données bactériennes, et ce avant toute analyse réalisée dans le cadre du travail de thèse. Cependant, aucune connaissance sur l'origine de ces contaminations, ni même leur concentration n'a pu être déduite de cette observation à l'œil nu lors des manipulations faites au laboratoire. Une autre étude semblait étayer cette observation en soupçonnant *H. ostrearia* de vivre avec sa propre "communauté bactérienne" (Lépinay et al 2018), mais toujours sans qu'aucune connaissance ne puisse être apportée sur leur concentration ou leur identité.

2. Aperçu général des données ONT et Illumina ainsi que du métagénome obtenu

Les paragraphes suivants détaillent les observations réalisées sur le métagénome obtenu en premier lieu avec les données ONT. La méthodologie utilisée pour la culture des *H. ostrearia* et pour la reconstruction de ce métagénome est la même que celle décrite dans l'article suivant (Chapitre 1-3), à la seule différence que j'ai utilisé ici des lectures longues ONT. Ce sont ces analyses qui ont motivé, notamment, la demande du deuxième jeu de données en PacBio CLR en mettant en évidence le peu de séquences identifiées comme appartenant au génome nucléaire de *H. ostrearia* (Annexes - Tentatives approfondies de décontamination du métagénome).

Les données de séquençage Illumina ont été filtrées pour éliminer les lectures de basse qualité (<Q30), de petite taille (<75b) ainsi que tout reste d'adaptateurs grâce à trimmomatic 0.39 (Bolger et al 2014). La qualité de ces données nettoyées et leur composition ont été observées grâce aux logiciels FastQC 0.11.8 et MultiQC 1.7 (Andrews 2010, Ewels et al 2016). Ce rapport présente une courbe retraçant le pourcentage moyen en nucléotides G et C (%GC) des lectures et leur abondance relative dans les données (Figure 1-1). En principe, la courbe à l'aspect d'une gaussienne, mais elle était ici composée de deux sommets : le premier à environ 45%GC représentant 20% des séquences et le second à 60%GC constituant plus de 40% des séquences. De plus, sachant que les génomes des microalgues *P. tricornutum* et *T. pseudonana* sont composés à 53,7%GC et 47%GC, il a été suggéré que seules les 20% de séquences à 45%GC pouvaient potentiellement contenir des lectures de *H. ostrearia*. Devant cette première observation, notre hypothèse était que les données étaient composées d'au moins deux organismes avec des %GC différents, ce qui justifiait la reconstitution d'un métassemblage et non d'un assemblage classique.

Ce phénomène a été également vérifié grâce à GenomeScope 1.0 (Vurture et al 2017) en partant des 21-mers produits par Jellyfish 2.2.3 (Marcais et Kingsford 2011) et contenus

dans les données de séquençage Illumina. Ce logiciel, développé pour estimer l'hétérozygotie d'un génome à partir d'un séquençage brut, a ici été employé pour vérifier si les deux populations observées précédemment pouvaient s'expliquer par une très grande variabilité dans les copies des gènes de *H. ostrearia*. Mais cette hypothèse a été mise à mal au vu de pics dans les données réelles dépassant le modèle analytique utilisé par GenomeScope pour modéliser ces dernières. En effet, la courbe noire du modèle s'alignait avec les données observées en bleu pour les premiers sommets de celles-ci (Figure 1-2, pics A et B). Les données étaient correctement modélisées, bien qu'il existait une petite divergence entre elles et leur modélisation vers 60 de "coverage" (profondeur). Cependant, deux autres pics pour les valeurs observées ont été remarqués après la "chute" de la courbe noire sur le deuxième graphique (Figure 1-2, pics C et D). De plus, selon la suggestion précédente basée sur le %GC que les données ne contiendraient qu'au maximum ~20% de *H. ostrearia*, il semblerait que les premiers pics A et B contiendraient des 21-mers appartenant au génome de la microalgue. Il y aurait donc énormément de contamination dans nos cultures par d'autres organismes (100X de profondeur pour plus de $1e^6$ 21-mers et plus de 40% des données Illumina). Le taux d'erreur indiqué par le logiciel était également en désaccord avec le taux d'erreurs associé à un séquençage Illumina : 0,9789% estimé contre 0,1 à 0,7% de probabilité d'erreur pour Illumina (Figure 1-2, Vilella 2023) ; ce qui peut être anormal et laisser encore penser à une contamination des données.

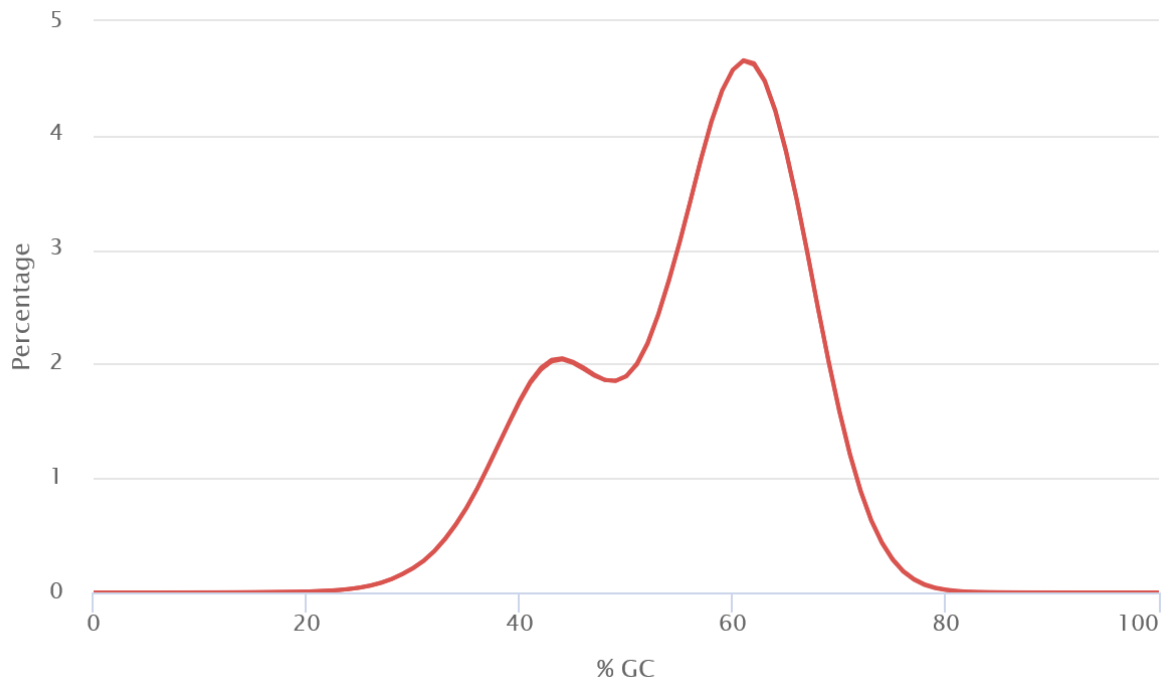


Figure 1-1 : Courbe modélisant le contenu en Guanine et Cytosine moyen (%GC) des lectures Illumina séquencées des cultures de *Haslea ostrearia*

Figure extraite du rapport obtenu grâce à MultiQC Report 1.7 (Ewels et al 2016) et FastQC 0.11.8 (Andrews 2010). Les lectures Illumina utilisées ont été préalablement nettoyées pour éliminer les lectures de basse qualité (<Q30), de petite taille (<75b) ainsi que tout reste d'adaptateurs grâce à trimmomatic 0.39 (Bolger et al 2014).

Results

GenomeScope version 1.0
k = 21

property	min	max
Heterozygosity	1.9143%	2.83622%
Genome Haploid Length	84,995,176 bp	86,625,830 bp
Genome Repeat Length	57,268,668 bp	58,367,382 bp
Genome Unique Length	27,726,508 bp	28,258,448 bp
Model Fit	67.3641%	88.5597%
Read Error Rate	0.97869%	0.97869%

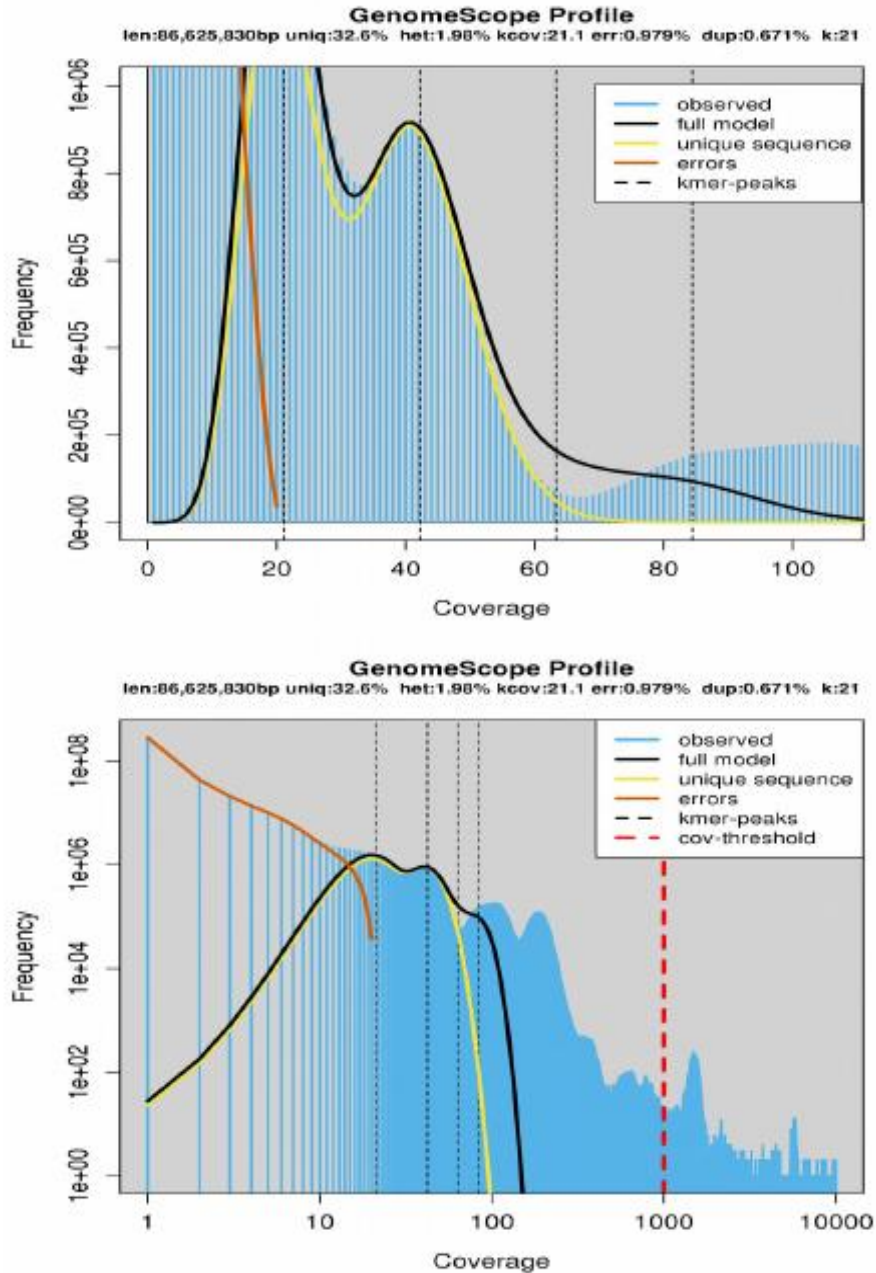


Figure 1-2 : Estimation de l'hétérozygotie, du contenu répété et de la taille des génomes contenus dans les lectures de séquençage Illumina des cultures de *Haslea ostrearia*

Figure réalisée à partir du logiciel GenomeScope 1.0 (Vurture et al 2017) utilisant une approche statistique basée sur les 21-mers présents dans les données ADN.

Les lectures Illumina utilisées ont été préalablement nettoyées pour éliminer les lectures de basse qualité (<Q30), de petite taille (<75b) ainsi que tout reste d'adaptateurs grâce à trimmomatic 0.39 (Bolger et al 2014).

Face à ces hypothèses, un métagénome *de novo* a été reconstruit grâce aux lectures longues, puis corrigé par les lectures courtes (Chapitre 1-3). Les caractéristiques physiques de ce métassemblage ont été obtenues par le logiciel Quast 5.0.2 (Gurevich et al 2013, Figure 1-3). Dans ces dernières, il a été mis en évidence un très grand nombre de contigs produits (2623) ainsi qu'une importante taille totale de ces derniers (>202 Gb). Pourtant, seulement 598 contigs ont été notifiés comme dépassant une taille de 50 kb, beaucoup n'atteignaient pas 25 kb (933 contigs, tableau de la Figure 1-3). De plus, lors des représentations graphiques du nombre de contigs / nombre de fenêtres glissantes par rapport au %GC, nous avons retrouvé les deux populations distinctes de séquences à 45%GC et 60%GC (graphiques de la Figure 1-3). Le grand nombre de contigs associés à 45%GC, mais leur plus faible nombre de fenêtres semblait également indiquer que les contigs étaient en plus grand nombre, mais également potentiellement plus courts que ceux associés à 60%GC. Il restait cependant à vérifier si les contigs à 45%GC appartenaient réellement à *H. ostrearia*, et si le %GC de son génome correspondait à nos attentes.

Face à cette contamination évidente des données, plusieurs méthodes ont été tentées pour extraire les séquences de *H. ostrearia* des données présentées ci-dessus. Cette décontamination du métagénome s'est cependant avérée hardue suite à la complexité des données séquencées (beaucoup de bactéries diverses, peu de profondeur de séquençage pour les contigs associés à *H. ostrearia*, etc.), et n'a pas produit les résultats escomptés. Afin de faciliter la lecture de ce manuscrit, les informations relatives aux différentes tentatives de décontamination sont disponibles dans les Annexes de celui-ci, dans la partie « Tentatives approfondies de décontamination du métagénome ».

HOB9_16_Sept2020_meta.flye.pilon	
# contigs (>= 0 bp)	2623
# contigs (>= 1000 bp)	2604
# contigs (>= 5000 bp)	2519
# contigs (>= 10000 bp)	2323
# contigs (>= 25000 bp)	1690
# contigs (>= 50000 bp)	598
Total length (>= 0 bp)	202933430
Total length (>= 1000 bp)	202920647
Total length (>= 5000 bp)	202597868
Total length (>= 10000 bp)	201110439
Total length (>= 25000 bp)	189051833
Total length (>= 50000 bp)	150713406
# contigs	2622
Largest contig	7116910
Total length	202933175
GC (%)	52.51
N50	273256
N75	48247
L50	63
L75	629
# N's per 100 kbp	0.30

All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

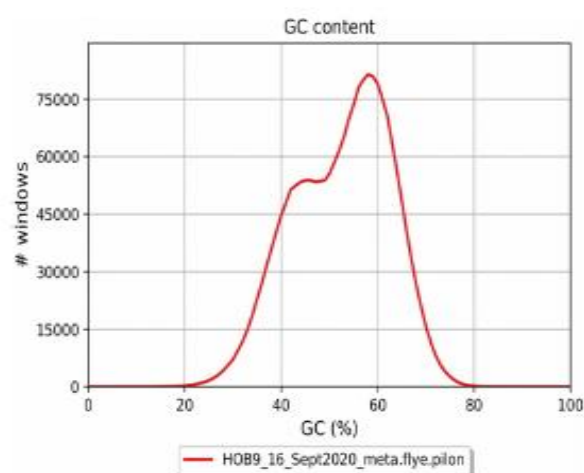
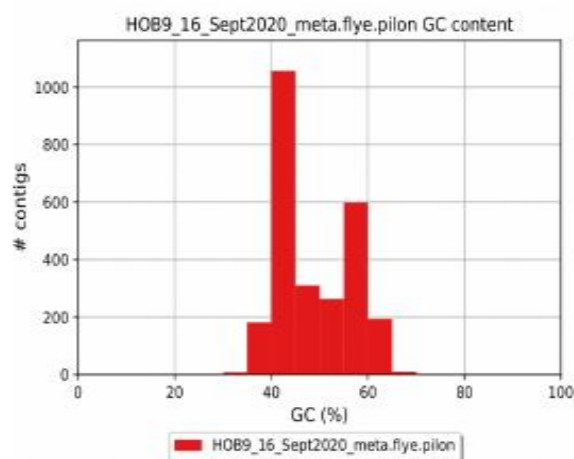


Figure 1-3 : Résumé des statistiques obtenus sur le métagénome obtenu à partir des séquençages ADN des cultures de *Haslea ostrearia*

Ces figures sont extraites du rapport Quast 5.0.2 (Gurevich et al 2013) sur l'assemblage réalisé par Flye 2.6 (Kolmogorov et al 2019) en mode métagénomique, et à partir des données de séquençage Nanopore et Illumina de Septembre 2020 des cultures de *Haslea ostrearia*.

3. Article sur les génomes chloroplastique et mitochondrial de

H. ostrearia

Face à la grande concentration de données bactériennes dans les séquençages (~80% des séquences brutes obtenues), un deuxième jeu de données a été demandé à la plateforme de séquençage Genotoul à Toulouse. Cette fois-ci l'ADN a été séquencé grâce à la technologie de PacBio CLR (PB, Tableau 1-1) afin d'obtenir un nombre de séquences plus important. L'objectif derrière cette demande était de contrebalancer l'immense contamination bactérienne par l'augmentation de la profondeur de séquençage, et ainsi augmenter le nombre de séquences appartenant à *H. ostrearia*. Cependant, suite à la trop grande diversité bactérienne présente dans les cultures (>30 espèces différentes), seuls le génome mitochondrial entier et celui quasi complet du chloroplaste ont pu être extraits des données. Les différentes techniques de décontamination citées précédemment n'ont toujours pas permis d'obtenir un nombre suffisant de contigs appartenant à la microalgue, bien que la profondeur associée au génome nucléaire est passée de ~10X à ~20X. Les deux génomes des organites de *H. ostrearia* sont décrits dans l'article suivant.

[Article publié chez *Mitochondrial DNA Part B: Resources*, DOI : 10.1080/23802359.2023.2268747]
[Les *Supplementary data* de l'article sont disponibles dans les Annexes]

1 **Complete mitochondrial genome and draft chloroplastic genome of *Haslea***
2 ***ostrearia* (Simonsen 1974)**

3

4 **Authors**

5 Aurelie Peticca^{a*}, Mostefa Fodil^a, Helene Gateau^a, Jean-Luc Mouget^a, Francois Sabot^b, Benoit Chenais^a,
6 Nathalie Casse^a

7 ^a BIOSSE (Biology of Organisms: Stress, Health, Environment), UFR Sciences et Techniques, Le Mans
8 Universite, F-72085 Le Mans, France

9 ^b DIADE, University of Montpellier, CIRAD, IRD, F-34394 Montpellier, France

10 *corresponding author: aurelie.peticca@univ-lemans.fr

11

12 **Abstract**

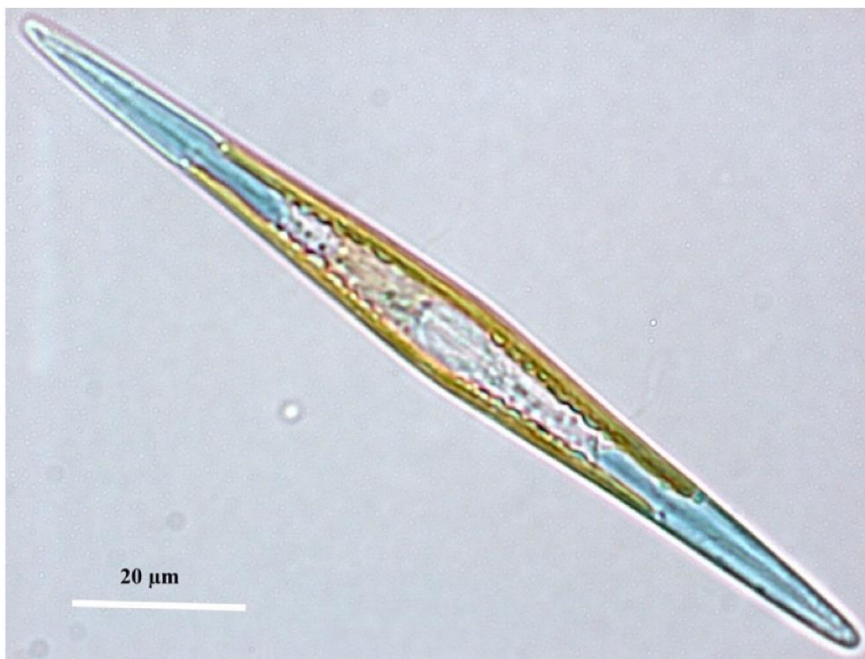
13 The first completed, circular mitochondrial genome and the first draft, linear chloroplastic genome of
14 the blue diatom *Haslea ostrearia* (Simonsen 1974, *Naviculaceae*, *Bacillariophyceae*) were assembled
15 from Illumina and PacBio sequencing. The mitochondrial genome was composed of 38,696 bases and
16 contained 64 genes, including 31 protein-coding genes (CDS), 2 ribosomal RNA (rRNA) genes and 23
17 transfer RNA (tRNA) genes. For the chloroplast, the genome was composed of 130,200 bases with 169
18 genes (131 CDS, 6 rRNA genes, 31 tRNA genes, and 1 transfer messenger RNA gene). Phylogenetic
19 trees, using the maximum-likelihood method and partial genes currently available for *Haslea ostrearia*
20 and other diatom species, suggested the proximity of all the *Haslea ostrearia* strains/isolates and the
21 possibility of using these genomes as future references.

22 **Keywords**

23 *Haslea ostrearia* ; diatom ; blue microalgae ; chloroplast ; mitochondrion

24 **Introduction**

25 *Haslea (H.) ostrearia* is a blue microalga from the Naviculaceae family, which lives freely in benthic
26 marine environments or as an epiphyte on brown macroalgae (Simonsen 1974, Figure 1). The blue
27 colour comes from a pigment called marenine, that *H. ostrearia* produces and accumulates at cell apices
28 (Gastineau et al. 2014a). This specific pigment is responsible for the greening of oyster gills in farming
29 ponds in Western France. Furthermore, it has been shown that marenine could display antibacterial,
30 antiviral and antifungal effects (Gastineau et al. 2014b). However, *H. ostrearia* is still very much
31 unknown, especially at the genetic level. One of the reasons for this lack of knowledge is that this
32 microalga needs many bacteria to survive (Lepinay et al 2016 and 2018). Their excessive presence
33 confuses the sequencing data. However, the complete DNA of one of its representatives from the North
34 Atlantic Ocean has been sequenced, and allowed the reconstruction of its mitochondrial and
35 chloroplastic genomes. This study clears the way for future research about *H. ostrearia* and will help
36 characterise the largely unknown genetics of this species. In particular, it will serve as a future basis for
37 the taxonomic classification of this species, but also as a potential marker for finding the presence of *H.*
38 *ostrearia* in metagenomic sequencing.



39

40 **Figure 1 : “Living cell of *Haslea ostrearia* observed in light microscopy” from Gabel et al (2022)**

41 **Material**

42 The isolate used in this study was collected in 2018 from an oyster pond in Bouin, France (latitude
43 46.953444 and longitude -2.046139) and deposited in the Nantes Culture Collection (curator Vona
44 Meleder, vona.meleder@univ-nantes.fr ; Nantes, France) under the name NCC 532.

45

46 **Methods**

47 During 21 days, the culture was grown in enriched artificial sea water (*Instant Ocean*, Aquarium systems
48 O ; Harrison et al. 1980 modified by De Brouwer et al. 2002) at 14°C under 300 µm photons/m²/s with
49 a 14h/10h light/dark cycle. On the 20th day of growth, a 1:100 dose of Sigma's antibiotic antimycotic
50 solution (Sigma-Aldrich, Saint-Quentin Fallavier, France; catalogue#A5955) was added to the culture
51 mix. After 24h, the biomass was collected through filtration (Whatman™ Binder-Free Glass Microfiber
52 Filters, Grade GF/C, pore size of 1.2 µm) and the whole DNA was extracted using the method of Puppo
53 et al. (2017). Extracted DNA was sequenced using the PacBio continuous long reads (PB CLR
54 SEQUEL2) and Illumina MiSeq platforms (TrueSeqv3, 150pe; Genotoul, Toulouse, France), and 85 Gb
55 and 5.241 Gb total read lengths were generated, respectively. Illumina reads were filtered to remove
56 low-quality reads (<Q30), short reads (<75b) and adapter sequences were searched and trimmed by
57 trimmomatic v0.39 (Bolger et al. 2014). The *de novo* genome assembly was performed using Flye v2.9
58 (Kolmogorov et al. 2020) and the long PB CLR reads with the following options: -g 100m --meta.
59 Polishing was performed with filtered Illumina reads through three loops of racon v1.4.20 (Vaser et al.
60 2017) and bwa-mem v2 2.2.1 (Vasimuddin et al. 2019). As the sequencing data also contained bacteria,
61 due to the characteristics of this microalgae which did not seem to be able to survive without them
62 (Lepinay et al 2016 and 2018). This is also the reason for the decision to sequence the data in long and
63 short reads, as it would have been difficult to obtain a quality assembly with only one of the latter (Chen
64 et al 2022). The mitochondrial and chloroplast genomes were identified using minimap2 v2.18 (Li 2018)
65 by aligning the metagenome obtained here against the *H. nusanlara* mitochondrion (MH681882
66 accession number, NCBI database, <https://www.ncbi.nlm.nih.gov/>) and the *Phaeodactylum tricornutum*

67 chloroplast (NC_008588, NCBI database). The sequences aligned with a percent identity superior to 80
68 were retained (%ID). The annotation and gene prediction were performed by Prokka v1.14.6 (Seeman
69 2014). Since the mitochondrial genome is circular, samtools faidx v1.14 (Danecek et al. 2021) was used
70 to relocate the annotated origin of H-strand replication (OH) as the starting gene (position +1). Attempts
71 to make the chloroplast circular using Circlator v1.5.5 (Hunt et al 2015) were unsuccessful, as the
72 sequences may be too dense at some point. Downsampling seems to be the solution to prevent an
73 assembly from failing due to information overload (Mirebrahim et al. 2015), but unfortunately with the
74 over-representation of bacteria this method cannot be applied. The supplementary figure 1 was made
75 according to the “Generating Sequencing Depth and Coverage Map for Organelle Genomes” in
76 protocol.io (<https://www.protocols.io/view/generating-sequencing-depth-and-coverage-map-for-o-4r3l27jkxg1y/v1>) using the assembled genome, the PacBio reads, minimap2 2.18 (Li et al. 2018) and
77 Samtools 1.14 (Danecek et al. 2021). Genome maps were generated with Artemis v18.2.0 (Rutherford
78 et al. 2000).

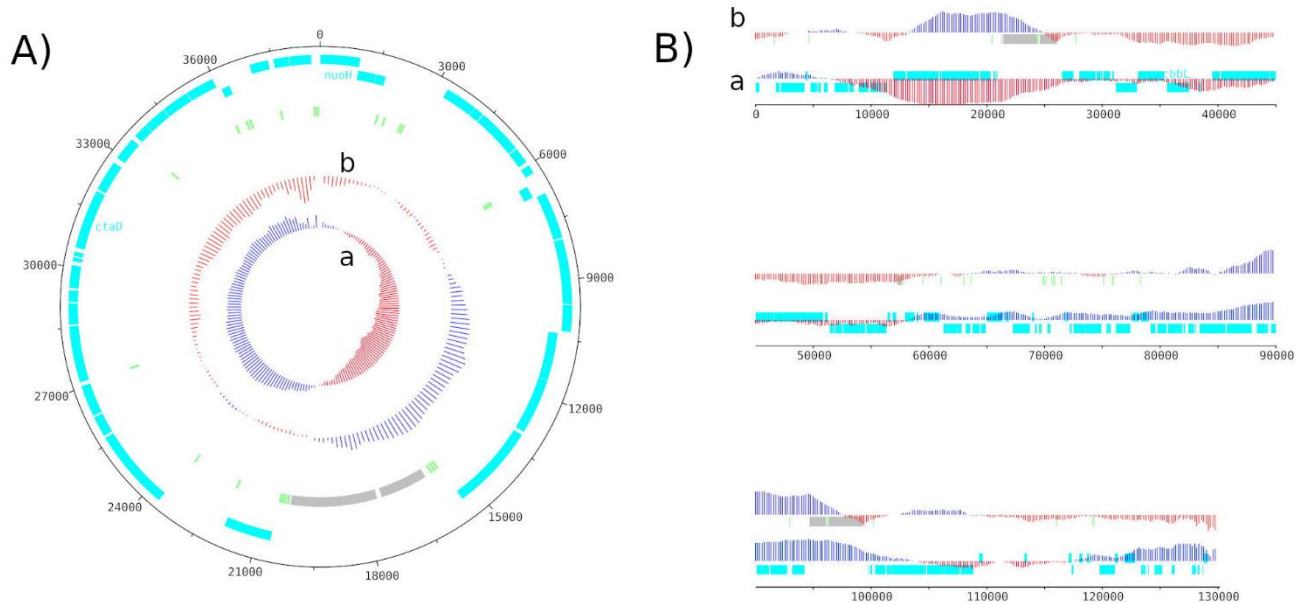
80 Phylogenetic trees were created with NGPhylogeny.fr pipeline (trimAl and PhyML+SMS, Lemoine et
81 al. 2019) with the partial *COX1* gene for the mitochondrial and the partial *rbcL* gene chloroplast
82 genomes, as these are the only data available for this species. The mitochondrial dataset grouped
83 sequences from different diatom taxa: 5 *H. ostrearia* strains, 4 other *Haslea* species, 5 *Navicula* species
84 and 2 external species from the *Eunotia* family. The chloroplast dataset also included different diatom
85 sequences: 3 *H. ostrearia* strains, 5 other *Haslea* species, 8 *Navicula* species and 2 external species from
86 the *Eunotia* family. All sequences were downloaded from NCBI database
87 (<https://www.ncbi.nlm.nih.gov/>), and their access number available on the phylogenetic trees.

88

89 **Results**

90 The complete circular mitochondrial genome was 38,696 bases long with a GC content (%GC) of
91 28.66% (36.07% A, 35.26% T, 14.76% C, 13.91% G), with a sequencing depth of 1,325X. The 64
92 annotated genes were composed of 39 protein-coding genes (CDS), 2 ribosomal RNA (rRNA) genes
93 and 23 transfer RNA (tRNA) genes (Figure 2A). The *H. ostrearia COX1* gene, previously partially

94 sequenced (Gastineau et al. 2013), was found complete in this study and annotated as *ctaD* by Prokka.
95 The draft chloroplast genome was 130,200 bases long with 31.04 %GC (34.19 A, 34.76% T, 15.31% C,
96 15.73% G) and with a sequencing depth of 2,371X. One hundred and thirty one CDS, 6 rRNA genes,
97 31 tRNA genes and 1 tmRNA gene were identified within this genome (Figure 2B).
98 The phylogenetic tree obtained for the *rbcL* gene from diatoms chloroplastic genomes showed a clear
99 separation between *Navicula* and *Haslea* species, with a bootstrap value of 0.818 (Figure 3A), the only
100 exception being the *rbcL* gene of *H. howeana* found among *Navicula*. Looking in details, the *H.*
101 *ostrearia* genes were clustered in the same part of the tree with very small branches. They were also
102 separated from the other diatoms by a branch with a bootstrap value of 0.986. The *Eunotia* genes, taken
103 as external species, were well observed on the external branches of the tree. The same observations were
104 made for the mitochondrial gene *COX1* (Figure 3B). *H. crucigera* was the only *Haslea* species found
105 the *Navicula*. The *COX1* genes from an *H. ostrearia* were all aggregated with small branches and were
106 separated from the others by a bootstrap value of 0.999. Due to a lack of information about genes in the
107 genus *Haslea*, only the partial *COX1* and the *rbcL* genes were tested.
108



109

110 **Figure 2 : Genome map of the mitochondrial genome (A) and the chloroplastic genome (B) of *H.***

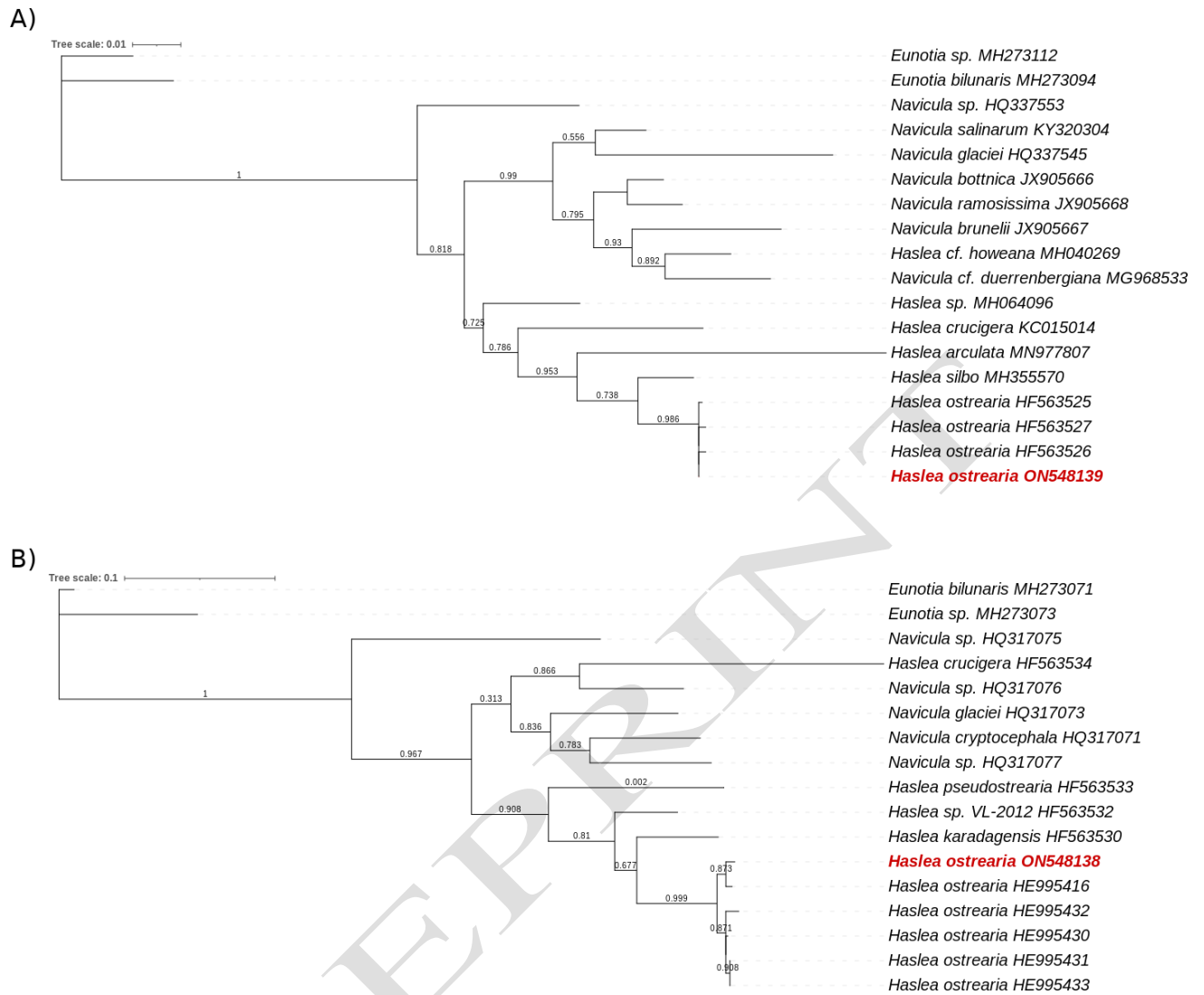
111 *ostrearia*

112 The red and blue colors in GC shiew (a) and GC plot (b) show if the value is below or above average.

113 Protein-coding genes are shown in light blue, transfer RNA genes in light green and ribosomal RNA in

114 light grey. The chloroplast genome is shown as a linear genome because it is not complete. The genome

115 map was made using Artemis v18.2.0.



116

117 **Figure 3 : Maximum likelihood phylogenetic trees inferred from COX1 and rbcL genes from**
 118 **diatoms genus**

119 The phylogenetic trees was performed with, respectevily, *Haslea ostrearia* (in red) and 16 to 18 other
 120 diatom chloroplastic gene *rbcL* (A) and mitochondrial gene *COX1* (B). Numbers near the nodes
 121 indicate bootstrap support values. The accession number associated with each gene is listed next to the
 122 species name. *Eunotia* species were used as external species. NGPhylogeny.fr pipeline was used to
 123 generate these phylogenetic trees.

124

125

126

127 Discussion and conclusion

128 Given the lack of information for the genus *Haslea*, the genomes reconstructed here could only be
129 compared directly with those of one other species to check their completeness. Both appear to be close
130 to the chloroplast and mitochondria of *H. nusantara*. The mitochondrion of the latter was 36,288 bases
131 long for a 29.24 %GC, and the chloroplast was 120,448 bases long for a 31.10 %GC (Prasetya et al.
132 2019), *i.e.* 2 kb and 10 kb more than those of *H. ostrearia*. The 64 annotated genes for the *H. ostrearia*
133 mitochondrion were slightly more numerous than *H. nusantara* ones with 3 additional CDS and one less
134 tRNA (Prasetya et al. 2019). Among the genes found here is *COXI*, which has been identified as being
135 highly conserved in the mitochondria of the *Haslea* (Gastineau et al. 2013). The mitochondria of *H.*
136 *ostrearia* appeared to be colinear with those of *H. nusantara*, the same synteny was observed in their
137 genomes in the form of three distinct blocks. The same comment could be made for the chloroplast, the
138 structure and the genes were very similar to the *H. nusantara* ones. The two chloroplast genomes closely
139 resembled each other, sharing close to 90% sequence identity. The only exceptions were a missing part
140 of the genome corresponding to ~10 kb of the *H. nusantara* chloroplast (positions around 90,000-
141 104,000) and the sequence inversions observed between *H. ostrearia* and *H. nusantara* for the first
142 70,000 bases of the latter. It was also very similar to the general features of diatom chloroplast genomes
143 from Prasetya et al. (2019) study. This amplified the idea that the genomes are complete for the
144 mitochondria and almost complete for the chloroplast, even if the latter has not been circularised. In
145 addition, the already known *H. ostrearia rbcL* gene, annotated as *cbbL*, and the *psbC* partial gene were
146 also found completed in this study (Gastineau et al. 2013). Phylogenetic trees of both mitochondrial and
147 chloroplastic sequences supported the hypothesis that *H. ostrearia* strains used here were very close to
148 each other and had a different evolutionary history from other diatoms or *Haslea* species (Figure 3).
149 High read coverage was also observed for each genome, respectively 1389X and 2321X for the
150 mitochondrial and the chloroplastic sequences (Supplementary Figure 1). In the case of the latter, there
151 is a short drop of up to 6X in coverage, but the average observed over the 4,000 bases it represents is
152 ~200X which is above the 30X traditionally desired for a *de novo* assembly.

153 Even if the chloroplast genome was not complete, it seemed that only a few bases were missing. Its size
154 exceeded that of the plastid from the close species *H. nusantara*, but with a highly similar composition
155 (%GC and genes number). Since all *H. ostrearia* strains, for which DNA sequences are available,
156 presented a close proximity to each other, the mitochondrial and chloroplast genomes presented here
157 could be used as a reference for this species.

158

159 **Ethical approval**

160 None required.

161

162 **Author contribution**

163 A.P, H.G, M.F were involved in the conception and design; H.G, M.F contributed to the DNA extraction;
164 A.P, F.S performed the analysis and interpretation of the data; A.P, F.S, B.C, N.C contributed the
165 drafting of the paper; H.G, M.F, J.L.M revised it critically for intellectual content; N.C and J.L.M found
166 the financing. All authors were involved in the final approval of the version to be published. All authors
167 agree to be accountable for all aspects of the work.

168

169 **Disclosure statement**

170 No potential conflict of interest was reported by the authors.

171

172 **Acknowledgement**

173 We express our sincere thanks to the Genotoul platform (Toulouse, France) for all sequencing data.

174

175

176 **Funding**

177 This research was supported by a Ph.D grant for A.P co-funded by Region Pays de la Loire (Lancom
178 project) and Le Mans University. M.F. benefited from the Horizon 2020 Research and Innovation
179 Program GHANA (The Genus *Haslea*, New marine resources for blue biotechnology and Aquaculture,
180 grant agreement No 734708/GHANA/H2020-MSCA-RISE-2016).

181

182 **Data availability statement**

183 The genome sequence data that support the findings of this study are openly available in GenBank of
184 NCBI at <https://www.ncbi.nlm.nih.gov/> under the accession no. ON548138 and ON548139. The
185 associated BioProject, SRA and Bio-Sample numbers are PRJNA843895,
186 SRR19450090/SRR19450089, and SAMN28772203 respectively.

187

188 **References**

189 Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 'Trimmomatic: A Flexible Trimmer for Illumina
190 Sequence Data'. *Bioinformatics* 2014, 30(15), 2114–2120.
191 <https://doi.org/10.1093/bioinformatics/btu170>.

192 Chen, L. et al. (2022) 'Short- and long-read metagenomics expand individualized structural variations
193 in gut microbiomes', *Nature Communications*, 13(1), p. 3175. Available at:
194 <https://doi.org/10.1038/s41467-022-30857-9>.

195 Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard,
196 Andrew Whitwham, et al. 'Twelve Years of SAMtools and BCFtools'. *GigaScience* 2021, 10(2),
197 giab008. <https://doi.org/10.1093/gigascience/giab008>.

198 De Brouwer, J. F. C., K. Wolfstein, and L. J. Stal. 'Physical Characterization and Diel Dynamics of
199 Different Fractions of Extracellular Polysaccharides in an Axenic Culture of a Benthic Diatom'.

200 European Journal of Phycology 2002, 37(1), 37–44.
201 <https://doi.org/10.1017/S0967026201003419>.

202 Gabed, N., Verret, F., Peticca, A., Kryvoruchko, I., Gastineau, R., Bosson, O., Seveno, J., Davidovich,
203 O., Davidovich, N., Witkowski, A., Kristoffersen, J.B., Benali, A., Ioannou, E., Koutsaviti, A.,
204 Roussis, V., Gâteau, H., Phimmaha, S., Leignel, V., Badawi, M., Khiar, F., Francezon, N., Fodil,
205 M., Pasetto, P., Mouget, J.-L., 2022. What Was Old Is New Again: The Pennate Diatom *Haslea*
206 *ostrearia* (Gaillon) Simonsen in the Multi-Omic Age. *Marine Drugs* 20, 234.
207 <https://doi.org/10.3390/md20040234>

208 Gastineau, Romain, Vincent Leignel, Boris Jacquette, Yann Hardivillier, Angela Wulff, Pierre Gaudin,
209 Djamel Bendahmane, Nicolai A. Davidovich, Irena Kaczmarska, and Jean-Luc Mouget.
210 ‘Inheritance of Mitochondrial DNA in the Pennate Diatom *Haslea Ostrearia* (Naviculaceae)
211 during Auxosporulation Suggests a Uniparental Transmission’. *Protist* 2013, 164(3), 340–351.
212 <https://doi.org/10.1016/j.protis.2013.01.001>.

213 Gastineau, Romain, Nikolai Davidovich, Gert Hansen, Jan Rines, Angela Wulff, Irena Kaczmarska,
214 James Ehrman, et al. ‘*Haslea Ostrearia*-like Diatoms’. In: *Sea Plants*, Bourgougnon N., editor.
215 *Advances in Botanical Research*, Academic Press, Vol. 71 t, p. 441–465, 2014a.
216 <https://doi.org/10.1016/B978-0-12-408062-1.00015-9>.

217 Gastineau, Romain, Francois Turcotte, Jean-Bernard Pouvreau, Michele Morançais, Joël Fleurence, Eko
218 Windarto, Fiddy Prasetya, et al. ‘Marennine, Promising Blue Pigments from a Widespread
219 *Haslea* Diatom Species Complex’. *Marine Drugs* 2014b, 12(6), 3161–3189.
220 <https://doi.org/10.3390/md12063161>.

221 Harrison, Paul J., Rosemary E. Waters, and F. J. R. Taylor. ‘A BROAD SPECTRUM ARTIFICIAL
222 SEA WATER MEDIUM FOR COASTAL AND OPEN OCEAN PHYTOPLANKTON 1’.
223 *Journal of Phycology* 1980, 16(1), 28–35. <https://doi.org/10.1111/j.0022-3646.1980.00028.x>.

224 Hunt, M. et al. (2015) ‘Circlator: automated circularization of genome assemblies using long sequencing
225 reads’, *Genome Biology*, 16(1), p. 294. Available at: [https://doi.org/10.1186/s13059-015-0849-](https://doi.org/10.1186/s13059-015-0849-0)
226 0.

227 Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong
228 Shin, Kristen Kuhn, et al. ‘MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat
229 Graphs’. *Nature Methods* 2020, 17(11), 1103–1110. [https://doi.org/10.1038/s41592-020-00971-](https://doi.org/10.1038/s41592-020-00971-x)
230 x.

231 Lemoine, Frederic, Damien Correia, Vincent Lefort, Olivia Doppelt-Azeroual, Fabien Mareuil, Sarah
232 Cohen-Boulakia, and Olivier Gascuel. ‘NGPhylogeny.Fr: New Generation Phylogenetic Services
233 for Non-Specialists’. *Nucleic Acids Research* 2019, 47(W1), W260–W265.
234 <https://doi.org/10.1093/nar/gkz303>.

235 Lepinay, A., Capioux, H., Turpin, V., Mondeguer, F., Lebeau, T., 2016. Bacterial community structure
236 of the marine diatom *Haslea ostrearia*. *Algal Research* 16, 418–426.
237 <https://doi.org/10.1016/j.algal.2016.04.011>

238 Lepinay, A., Turpin, V., Mondeguer, F., Grandet-Marchant, Q., Capioux, H., Baron, R., Lebeau, T.,
239 2018. First insight on interactions between bacteria and the marine diatom *Haslea ostrearia*: Algal
240 growth and metabolomic fingerprinting. *Algal Research* 31, 395–405.
241 <https://doi.org/10.1016/j.algal.2018.02.023>

242 Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100.
243 <https://doi.org/10.1093/bioinformatics/bty191>

244 Mirebrahim, H., Close, T.J. and Lonardi, S. (2015) ‘De novo meta-assembly of ultra-deep sequencing
245 data’, *Bioinformatics*, 31(12), pp. i9–i16. Available at:
246 <https://doi.org/10.1093/bioinformatics/btv226>.

247 Prasetya, Fiddy Semba, Romain Gastineau, Michel Poulin, Claude Lemieux, Monique Turmel, Agung
248 Dhamar Syakti, Yann Hardivillier, et al. ‘*Haslea Nusantara* (Bacillariophyceae), a New Blue

249 Diatom from the Java Sea, Indonesia: Morphology, Biometry and Molecular Characterization'.
250 Plant Ecology and Evolution 2019, 152(2), 188–202. <https://doi.org/10.5091/plecevo.2019.1623>.

251 Puppo, Carine, Tomas Voisin, and Brigitte Gontero. 'Genomic DNA Extraction from the Pennate
252 Diatom Asterionella Formosa Optimised for next Generation Sequencing. V1'. Preprint, 21
253 September 2017. <https://doi.org/10.17504/protocols.io.jytcpxn>.

254 Seemann, T. 'Prokka: Rapid Prokaryotic Genome Annotation'. Bioinformatics 2014, 30(14), 2068–
255 2069. <https://doi.org/10.1093/bioinformatics/btu153>.

256 Simonsen, R. The Diatom Plankton of the Indian Ocean Expedition of RV 'Meteor' 1964 - 1965.
257 'Meteor' Forschungsergebnisse Reihe D, Biologie, No. 19. Berlin Stuttgart: Borntraeger, 1974.

258 Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 'Fast and Accurate de Novo Genome
259 Assembly from Long Uncorrected Reads'. Genome Research 2017, 27(5), 737–746.
260 <https://doi.org/10.1101/gr.214270.116>.

261 Vasimuddin, Md., Sanchit Misra, Heng Li, and Srinivas Aluru. 'Efficient Architecture-Aware
262 Acceleration of BWA-MEM for Multicore Systems'. In 2019 IEEE International Parallel and
263 Distributed Processing Symposium (IPDPS), 314–24. Rio de Janeiro, Brazil: IEEE, 2019.
264 <https://doi.org/10.1109/IPDPS.2019.00041>.

4. Conclusion du chapitre et prélude du chapitre suivant

L'objectif initial était de reconstituer le génome de *H. ostrearia* pour ensuite pouvoir étudier sa composition durant la suite de la thèse. Cependant, une forte contamination des données par de nombreux génomes bactériens a été mise en évidence et ce même lors de l'étude des données de séquençage brutes. Même après de nombreuses tentatives pour décontaminer le génome, la quantité de données récupérée sur *H. ostrearia* n'a pas été assez importante pour l'étude de sa composition. Un séquençage plus profond ne résolvant pas le problème, seuls les génomes entiers de la mitochondrie et presque entiers du chloroplaste ont pu être extraits des données.

Devant cet obstacle, il a été décidé de réorienter la problématique initiale et de se concentrer sur les informations contenues dans le dernier métagénome obtenu (PB+Illumina). En effet, beaucoup d'informations sur les bactéries présentes dans les cultures de *H. ostrearia* nous sont offertes par ce dernier : il contient de nombreux génomes bactériens reconstruits et entiers ou quasiment. Bien que *H. ostrearia* ait déjà été soupçonnée de vivre avec une communauté bactérienne (Lépinay et al 2018), l'identification de ces bactéries n'a jamais été proposée auparavant. De plus, connaître ces bactéries permettrait d'améliorer le protocole de cultures actuel en réduisant leur concentration ou leur diversité. Actuellement la population bactérienne présente dans celles-ci n'est pas prise en compte par le protocole, ce qui impacte négativement la reproductivité des cultures et l'accessibilité du génome de *H. ostrearia* dans les séquençages. Afin d'axéniser ou de purifier au maximum les cultures, il est nécessaire de connaître les bactéries présentes.

Chapitre 2 : Identifier les bactéries présentes dans les cultures

1. Contexte des travaux de recherche

Dans le but d'en apprendre plus sur les bactéries présentes dans les cultures de *H. ostrearia*, il a été proposé d'utiliser un séquençage 16S. Avec cette méthode conventionnelle pour l'identification bactérienne, même les bactéries rares en petites concentrations dans les cultures ont pu être détectées. Ceci présentait l'avantage de pouvoir vérifier que tous les microorganismes retrouvés provenaient bien d'un environnement marin, et donc qu'il n'y avait pas eu de contamination des cultures : ils auraient été isolés en même temps que la microalgue. Également, avec cette méthode, il a été possible d'estimer la diversité des bactéries associées aux cultures, ce qui était très intéressant puisque beaucoup de bactéries différentes avaient été retrouvées dans les métagénomomes et ce même en ne considérant que l'identification taxonomique des contigs les mieux assemblés (>1Mb). Il y avait donc une grande diversité dans les taxons bactériens identifiés.

Aussi, pour aller dans le sens du nouvel objectif d'amélioration du protocole de culture, cette identification bactérienne a été réalisée pour deux protocoles de culture différents. Celui habituellement utilisé en laboratoire, qui ne contenait aucune étape pour réduire le nombre de bactéries dans les cultures, et un deuxième qui reprenait les étapes du premier en ajoutant un traitement léger des cultures par des antibiotiques à large spectre. Les résultats ainsi obtenus ont permis d'estimer l'effet de cette modification du protocole de culture sur la population bactérienne, que cela soit au niveau de la concentration ou au niveau de la diversité des espèces. Ce traitement a également été qualifié de "léger" puisque les doses d'antibiotique administrées étaient réduites, sur une courte période, pour s'assurer de la survie de *H. ostrearia* dans les cultures. En effet, il avait été précédemment remarqué au laboratoire qu'une dose trop importante ou une exposition longue avait tendance à tuer la microalgue. Aucune explication n'a été donnée à ce jour à ces observations.

L'article suivant retrace les observations et hypothèses émises durant l'identification bactérienne par un séquençage 16S et ce pour 2 types de cultures différentes.

2. Article sur l'identification par 16S des bactéries présentes dans les cultures de *H. ostrearia*

[Article en cours de relecture chez *Scientific Data*]
[Les *Supplementary data* de l'article sont disponibles dans les Annexes]

1 **To improve the axenic culture protocol, the bacteria present in *Haslea ostrearia***
2 **cultures need to be studied**

3 Aurelie Peticca^a, Mostefa Fodil^a, Benoit Chenais^{a+}, Francois Sabot^{b+*} and Nathalie Casse^{a+}

4 *Corresponding author, francois.sabot@ird.fr

5 ⁺Co-last authors

6 ^aBiOSSE, UFR Sciences et Techniques, University of Le Mans, F-72085 Le Mans, France

7 ^bDIADE, University of Montpellier, CIRAD, IRD, F-34394 Montpellier, France

8 **ORCID numbers :**

9 AP [0000-0002-8972-0973](https://orcid.org/0000-0002-8972-0973)

10 MF [0000-0003-2856-1988](https://orcid.org/0000-0003-2856-1988)

11 BC [0000-0003-2441-1966](https://orcid.org/0000-0003-2441-1966)

12 FS [0000-0002-8522-7583](https://orcid.org/0000-0002-8522-7583)

13 NC [0000-0003-4663-8491](https://orcid.org/0000-0003-4663-8491)

14

15 **Abstract**

16 **Objectives:** *Haslea ostrearia* is a microalgae associated with marenine, a blue pigment known for its
17 economic interest when it attaches to oyster gills. Complex to grow under axenic conditions, *H. ostrearia*
18 probably needs its own bacterial consortium to survive. Consequently, this hypothetical important
19 bacterial presence greatly complicates any study of its ecological, physiological and genomic content.
20 Identifying and monitoring the concentration of these bacteria would allow the current cultivation
21 protocol to be optimised in order to improve access to information on *Haslea ostrearia*.

22 **Results:** We therefore tested a first approach of a protocol to control the bacterial concentration without
23 killing the microalgae, by growing four *H. ostrearia* NCC 532 cultures under two culture conditions:
24 two with and two without antibiotic treatment. For the first time, a bacterial identification technique
25 using 16S rRNA sequencing was applied to the bacterial community of *H. ostrearia* cultures. The V3V4
26 16S region allowed precise identification of these bacteria. To ensure a non-variant bacterial population
27 in the cultures prior to DNA extraction, antibiotic treatment was also deemed necessary.

28 **Keywords:** 16S, reproducibility, diatom, cultures, treatment, *Haslea ostrearia*, bacterial community

30 Introduction

31 *Haslea ostrearia* is a microalgae species from the diatom family (*Bacillariophyta*) [1], highly present in
32 the oyster beds of the French Atlantic coast [2]. Its peculiarity lies in its bluish colour, caused by a
33 pigment called marenine [3], which has made *H. ostrearia* a subject of study for its economic interest,
34 but also for its possible biomedical applications [4, 5]. In fact, the strong presence of *H. ostrearia* in
35 oyster cultures gives a greenish colour to the gills of the oysters [6, 7, 8]. Beyond this physical aspect,
36 it appears that its presence in adult oysters significantly reduces their oxygen consumption, filtration
37 and feeding activity [8, 9, 10]. The increasing market value of these colourful oysters in France is due
38 to their rarity and beauty.

39 Therefore, a better knowledge of *H. ostrearia* would allow us to understand how marenine is produced
40 and its potential uses. Unfortunately, due to its ecological complexity, *H. ostrearia* is difficult to grow
41 in axenic conditions, as it seems to live with its own mandatory bacterial community [11, 12]. However,
42 the high and variable concentration of bacteria in the data makes access to the *H. ostrearia* genome, and
43 therefore any information about it, difficult. To the best of our knowledge, the identification of the
44 bacterial taxa present in the cultures has never been studied before and seems to us to be necessary in
45 order to obtain more specific culture conditions.

46 Therefore, this study aims to propose a first identification of the bacteria associated with *H. ostrearia*
47 and their biocontrol. By controlling this bacterial population, it would be possible to optimise the
48 cultivation of these microalgae and ensure the reproducibility of the results, i.e. to obtain the same
49 bacterial population with each new cultivation. The data used were from four cultures of the same *H.*
50 *ostrearia* isolate (NCC 532) but under two different conditions. For two of them, broad-spectrum
51 antibiotics were added to the usual laboratory conditions. For the other two, the culture protocol was
52 kept the same as it is normally used. The whole DNA of each culture was then independently extracted.
53 The V1V3 and V3V4 regions of 16S ribosomal RNA were amplified and sequenced using Illumina
54 technology. To determine which region best identifies *H. ostrearia* associated bacteria, a Mock
55 community sequencing was also performed and compared to the expected results.

56

57 **Main text**

58 ***Haslea ostrearia* cultures and DNA extraction**

59 The selected isolate for this study was recovered the 16 January 2018 from an oyster pond in Bouin,
60 France (46°57'12.4"N, 2°02'46.1"W). Four cultures were performed according to the usual protocol for
61 these microalga: Harrison et al. 1980 [13] modified by De Brouwer et al. 2002 [14] (Supplementary
62 data). After 20 days, two of the four cultures received a 1:100 dose of antimycotic antibiotic solution
63 (Sigma-Aldrich, catalogue #A5955) containing penicillin, streptomycin and amphotericin B. The use of
64 broad-spectrum antibiotics was justified by the lack of knowledge about the bacteria present in the
65 cultures. The other two were left untreated to serve as a control for the culture protocol currently in use.
66 Total DNA was extracted on day 21 of growth, i.e. the plateau of the culture, and sent to the Genotoul
67 platform (Toulouse, France) for the subsequent protocol. The total DNA was then amplified,
68 independently, using two different sets of primer designed for two 16S regions. They were next
69 sequenced by Illumina MiSeq platforms (TrueSeqv3, 250pe) with a negative control sample without any

70 organisms (Supplementary data) and the standard ZymoBIOMICS® Microbial Community control
71 (D6300, ZymoResearch).

72

73 **Data availability, treatment and analyzes**

74 The scripts are available on GitHub: https://github.com/Thrylia/16S_Haslea_ostrearia [15]. The
75 sequencing obtained for *H. ostrearia* samples on the NCBI database under the BioProject
76 PRJNA873049 [16]. Supplementary data contains all information on the analysis protocol.

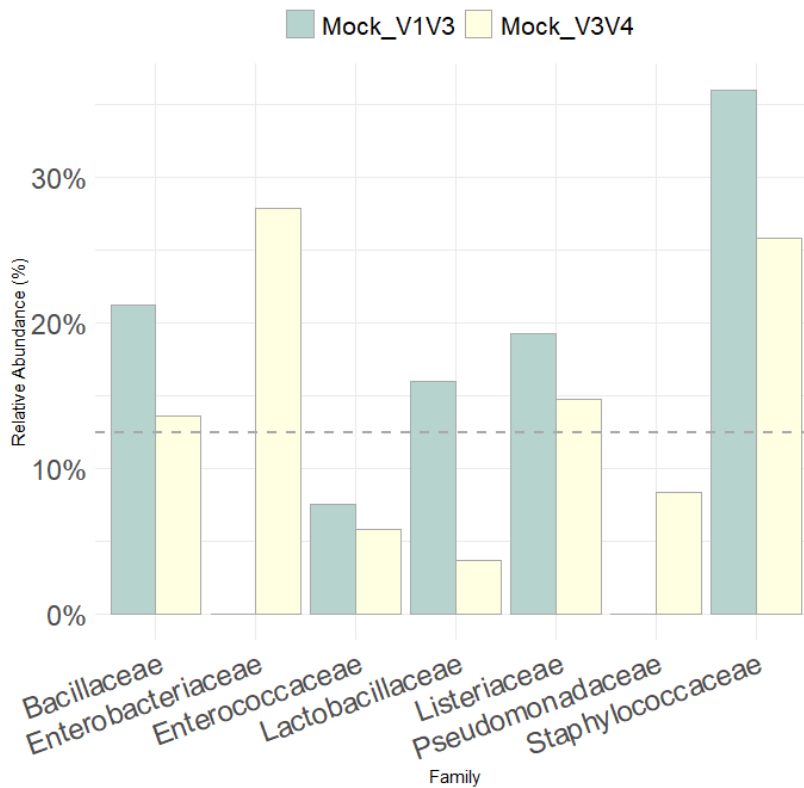
77

78 **Mock community sequencing**

79 In order to determine the best method to identify these bacteria, a known and homogeneous bacterial
80 community (Mock) was sequenced. This Mock community was constituted by 8 bacteria and 2 yeast
81 clades, in the same proportions (Fig. 1). The bacteria section was composed of *Pseudomonas*
82 *aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Lactobacillus faecalis*,
83 *Staphylococcus aureus*, *Listeria monocytogenes*, and *Bacillus subtilis*. However, only the bacteria can
84 be identified by the 16S method, and the selection of the 16S region is of great importance for bacterial
85 identification [17]. This mock community was therefore sequenced using two 16S control regions:
86 V1V3 and V3V4, which are widely used for bacterial identification [17]. V3V4 sequencing recovered
87 all the bacteria of the Mock community, which was not the case for V1V3 with which three of them
88 were missing: *P. aeruginosa*, *E. coli* and *S. enterica* (Fig. 1). No *Gammaproteobacteria* appeared in the
89 V1V3 results. As for the relative abundance, the observed values were slightly different from the 12.5%
90 expected. Yet, no taxon overwhelmingly dominated the others. V3V4 sequencing appears to correctly
91 detect genera present in a diverse bacterial community and provide a good overview of their abundance
92 in samples. Therefore, it was decided to continue with the V3V4 method for sequencing *H. ostrearia*
93 samples.

94 To identify any external contamination of the samples a negative control was also performed
95 (Supplementary data). This contained nothing but pure water and the reagents that were needed for the
96 PCR and the sequencing protocol.

97



98

99 **Figure 1:** Grouped barplot showing the relative abundance of the bacteria (Family) identified by
100 the 16S regions V1V3 (green) and V3V4 (yellow) in the Mock community samples.

101 The dotted line represents a relative abundance of 12.5% which corresponds to the abundance of each
102 bacterial species contained in the Mock community. The *Enterobacteriaceae* family regroups the
103 species *Escherichia coli* and *Salmonella enterica*.

104

105

106

107 **The variation of the bacterial population of *H. ostrearia* NCC 532**

108 A total of 156 bacterial taxa, including 35 orders, were found when considering V3V4 *H. ostrearia*
109 samples (Additional Table 2).

110 When grouped by class, peculiarities were noticed for bacteria from the V3V4-non treated samples (N
111 samples - Fig. 2, Additional Table 2): i) a very low abundance of *Oxyphotobacteria* and
112 *Deltaproteobacteria* in N1 sample; and ii) a large variation between the two non-treated samples (N1
113 and N2) for the relative abundance of *Alphaproteobacteria* and *Clostridia*, by almost 10 and 4 points
114 respectively. These differences between N1 and N2 intensified when orders were considered. For
115 instance, *Gammaproteobacteria* was composed of a majority of *Oceanospirillales* for N1, but this class
116 was shared between *Alteromonadales*, *Cellvibrionales* and *Oceanospirillales* for N2. The overall and
117 detailed profiles of the V3V4-N samples appeared to vary greatly between the two samples (Fig. 2, 3).

118 In an opposite way, no such discrepancy was found for V3V4-treated samples. The overall profile was
119 quite similar for the two samples with a high abundance of *Alphaproteobacteria* ($63\pm 1\%$),
120 *Gammaproteobacteria* ($\sim 17.8\pm 0.3\%$) and *Bacteroidia* ($\sim 11.5\pm 0.5\%$, Fig. 2). Only the low presence
121 of *Oxyphotobacteria* in the first treated-sample (T1) seemed to diverge from the second T2, but for less
122 than 1 point (Additional Table 2). The V3V4-T samples profiles also looked very similar for the detailed
123 profile (orders, Fig. 3). No variation between the results obtained for T1 and T2 above the arbitrary
124 threshold of 1% was observed. The samples were mainly composed of *Rhodobacterales*, *Rhizobiales*
125 and *Caulobacterales* (*Alphaproteobacteria*), representing $\sim 39\pm 0.2\%$, $\sim 13.5\pm 0.2\%$ and $\sim 6\pm 0.4\%$
126 respectively, of $\sim 9.8\pm 0.5\%$ of *Cellvibrionales* (*Gammaproteobacteria*) and of $\sim 8.5\%$ of
127 *Flavobacteriales* (*Bacteroidia*). To validate this observation, a χ^2 test was performed with the null
128 hypothesis that the samples were from the same population. The test result was significant and showed
129 that T samples shared the same composition of bacteria (p -value= $2.2e-16$). It is therefore suggested that
130 the addition of an antibiotic treatment can help conserve the same bacterial population with each new
131 culture of *H. ostrearia*.

132

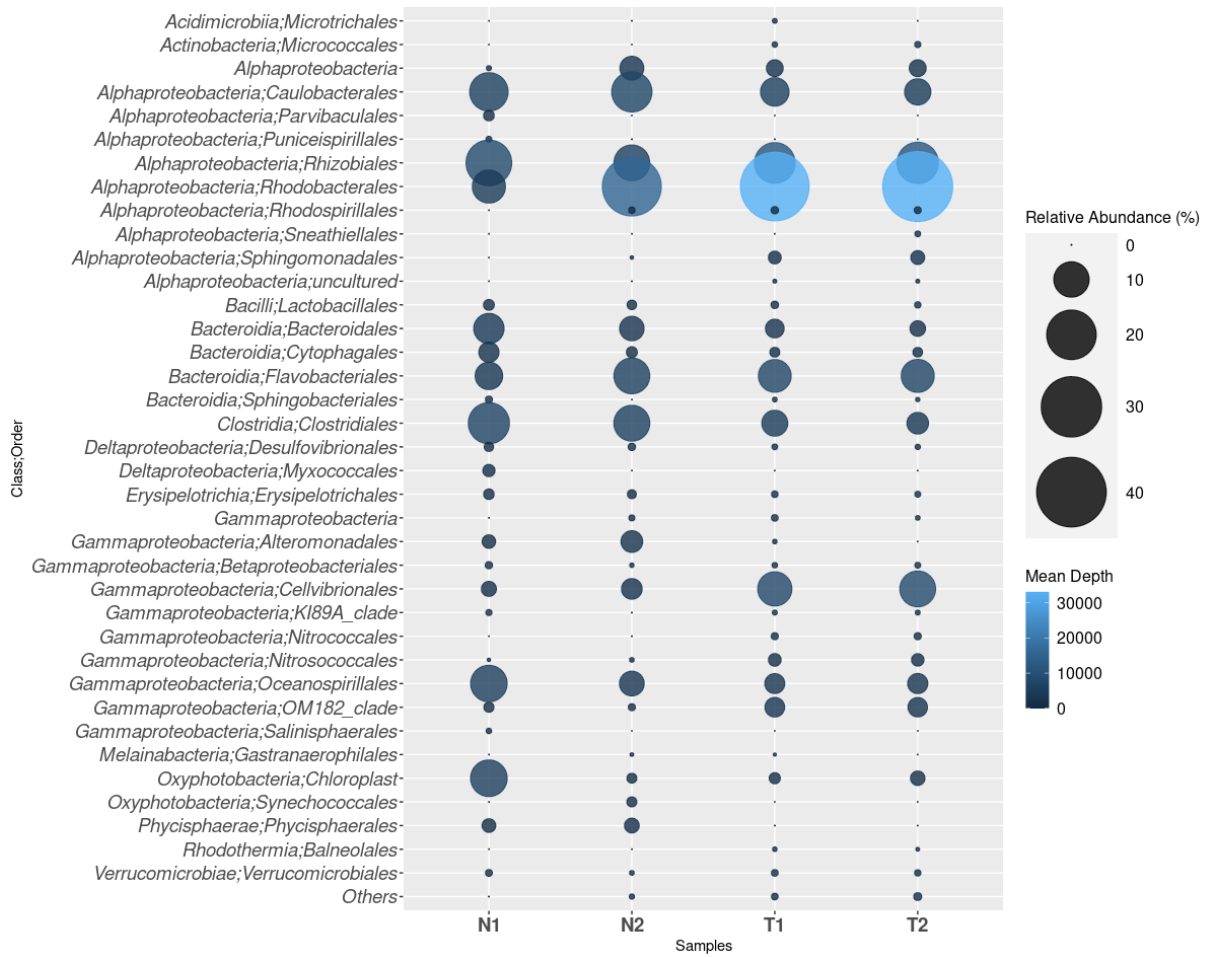


133

134 **Figure 2:** Stacked barplot showing the relative abundance of the bacteria (Phylum;Class) for the
 135 different samples of *Haslea ostrearia* identified by the 16S region V3V4.

136 Only classes with a relative abundance greater than or equal to 1% of the sample are represented on the
 137 graph. Classes not meeting this threshold have been grouped in the "Other" category. Samples beginning
 138 with an N represent the usual laboratory protocol for culturing *H. ostrearia*, whereas a T represents an
 139 addition of antibiotic to it.

140



142

143 **Figure 3: Bubblechart showing the relative abundance and the mean depth of the bacteria**
 144 **(Class;Order) in the samples N and T of *Haslea ostrearia* identified by the 16S region V3V4.**

145 Only orders with a relative abundance greater than or equal to 1% of the sample are represented on the
 146 graph. Orders not meeting this threshold have been grouped in the "Other" category. When a proportion
 147 of a class is not associated with any order, only the class is displayed. The mean depth was obtained
 148 using the ratio of the number of sequences aligned on each scaffold used for taxonomic annotation to
 149 the size of these scaffolds. Samples marked with an N correspond to the usual laboratory protocol for
 150 culturing *H. ostrearia*, whereas a T corresponds to the addition of antibiotics in the same laboratory
 151 protocol.

152

153 **The bacterial population of *H. ostrearia* NCC 532 compared to other diatoms**

154 For the reasons mentioned above, the T samples were used to describe the bacteria still remaining after
155 broad spectrum antibiotic treatment in the *H. ostrearia* cultures. More than 30 orders were found in the
156 latter (Fig. 3). The three most abundant phyla described here (Fig. 2) have also been found in other
157 studies of microalgae, including diatoms [18, 19]. *Proteobacteria*, specifically the *Alphaproteobacteria*,
158 were the majority of the bacteria found in microalgae's bacterial consortium, including the diatom
159 *Thalassiosira pseudonana* [19]. In addition, a previous study showed a high proportion of
160 *Gammaproteobacteria* in diatom cultures, which tended to increase with the culture time [18]. This
161 might explain why they accounted for about a quarter of V3V4-T samples, since the *H. ostrearia*
162 samples used here have been isolated in 2018 and have been grown discontinuously since. In this same
163 study, *Bacteroidetes* were also present in the cultures, as in *H. ostrearia* ones.

164 Literature on the effect of these bacteria on microalgae even led to consider the hypothesis that the
165 presence of these bacteria was not a mere coincidence [20, 21]. Indeed, here are the three most abundant
166 classes present in all *H. ostrearia* samples (~47-6%, Fig. 3): *Rhodobacterales* is a marine bacterium
167 suspected of having a growth enhancement properties for microalgae [20], *Rhizobiales* with the same
168 properties but for cultures of the green microalgae *Chlorella* [21], and *Flavobacteriales* that shown a
169 growth inhibitory effect to the microalgae *Nannochlorosis* sp. CCAP211/78 [22]. However, the specific
170 effect of these bacteria on *H. ostrearia* still needs to be described. For the other orders, especially for
171 the ones with a small abundance (<1%), it could be possible that the V3V4 region underestimates their
172 presence, or that they were indeed relatively rare in our samples. This question could be answered by
173 real-time quantitative PCR.

174 **Limitations**

175 Considering the high and diverse bacterial concentration in *H. ostrearia* cultures, it is undeniable that a
176 protocol to obtain a standardized bacterial population in each new culture is essential for future research.
177 Antibiotic treatment appears to be necessary to maintain the stability of the bacterial population, i.e. a
178 bacterial population that varies little or not at all with each new culture. However, to study *H. ostrearia*

179 DNA, current treatment with broad-spectrum antibiotics does not reduce the bacterial population
180 sufficiently. The microalgae remains at too low a concentration in the samples as shown by the low
181 abundance of the microalgae's chloroplast, which represented only 1% in the T samples. It is therefore
182 necessary to look for another antibiotic treatment, the one used in this study being broad spectrum due
183 to the lack of knowledge on bacteria before this analysis. More specific antibiotics to the bacteria
184 described here would increase the proportion of this microalgae in the cultures (and thus in the
185 sequencing data). A first strategy might be to use an antibiotic that targets *Flavobacteriales*, which
186 appear to inhibit microalgal growth [22] and are present in high concentrations in the cultures, but avoids
187 *Rhodobacteriales* and *Rhizobiales*, which appear to have growth-promoting properties for microalgae
188 [20, 21]. A second approach could be to test a range of antibiotics that target specific bacterial classes
189 and see if *H. ostrearia* survives and at what concentration. With this in mind, we are currently
190 investigating the antibiotic resistance genes (ARGs) present in the genomes of these bacteria by aligning
191 their metagenomic-assembled genomes against ARGs databases. However, as mentioned above, this
192 hypothetical antibiotic treatment does not have to kill all the bacteria, as some appear to be necessary
193 for *H. ostrearia* to grow and survive. The analysis of the impact of these antibiotics on the bacterial
194 population could be followed through sequencing of the V3V4 16S region, as well as the survival of *H.*
195 *ostrearia* without certain bacteria. This could enable us to determine the bacteria essential to the
196 microalgae environment. Finally, this research perspective needs to include a larger number of samples,
197 especially from other geographical areas. If the assumption that the essential bacteria of *H. ostrearia*
198 should always be found with the microalgae is correct.

199

200 **List of abbreviations**

- 201 - 16S: 16S ribosomal RNA
- 202 - NCC: Nantes Culture Collection
- 203 - N sample: Non treated sample

- 204 - Pe: Paired-end
- 205 - T sample: Treated sample (antibiotic treatment)
- 206 - T-neg sample: Negative control sample

207

208 **Ethics approval and consent to participate**

209 This study does not require any ethical or other relevant approvals.

210

211 **Consent for publication**

212 Not applicable.

213

214 **Availability of data and materials**

215 The dataset supporting the conclusions of this article is available in the NCBI database under the
216 BioProject PRJNA873049 at www.ncbi.nlm.nih.gov. All the scripts are available on GitHub
217 (https://github.com/Thrylia/16S_Haslea_ostrearia).

218

219 **Competing interests**

220 None.

221

222 **Funding**

223 This research was supported by the Region Pays de la Loire (Lancom project) and Ph.D. grant for A.P.
224 was co-funded by Region Pays de la Loire (Lancom project) and Le Mans University. M.F. benefited
225 from the Horizon 2020 Research and Innovation Program GHaNA (The Genus Haslea, New marine
226 resources for blue biotechnology and Aquaculture, grant agreement No 734708/GHANA/H2020-
227 MSCA-RISE-2016).

228

229 **Authors' contributions**

230 Conception and design, AP, FS, BC, NC; analysis and interpretation, AP, FS; providing biological
231 material and DNA, MF; writing-original draft preparation, AP; writing-review and editing, AP, FS, BC,
232 NC; supervision, FS, BC, NC; funding acquisition, NC.

233 All authors agree to be accountable for all aspects of the work.

234

235 **Acknowledgements**

236 We express our sincere thanks to M Bruno Cognie for isolating the strain NCC 532 studied here. Also
237 to Ms Helene Gateau who maintained the cultures of *Haslea ostrearia*, and M Jean-Luc Mouguet who
238 is at the origin of the GHaNA program.

239 All sequencing analyses have been carried out by the Genotoul platform (Toulouse, France).

240

241

242

243 **References**

- 244 1. R. Simonsen, *The Diatom Plankton of the Indian Ocean Expedition of RV 'Meteor' 1964-1965*.
245 Gebrüder Borntraeger, 1974.
- 246 2. R. Gastineau *et al.*, 'Haslea ostrearia-like Diatoms', in *Advances in Botanical Research*, vol. 71,
247 Elsevier, 2014, pp. 441–465. doi: [10.1016/B978-0-12-408062-1.00015-9](https://doi.org/10.1016/B978-0-12-408062-1.00015-9).
- 248 3. J.-B. Pouvreau *et al.*, 'Preliminary characterisation of the blue-green pigment "marenine" from the
249 marine tychopelagic diatom *Haslea ostrearia* (Gaillon/Bory) Simonsen', *J Appl Phycol*, vol. 18,
250 no. 6, pp. 757–767, Nov. 2006, doi: [10.1007/s10811-006-9087-x](https://doi.org/10.1007/s10811-006-9087-x).
- 251 4. J.-B. Pouvreau, F. Taran, P. Rosa, S. Pin, J. Fleurence, and P. Pondaven, 'Antioxidant and Free
252 Radical Scavenging Properties of Marenine, a Blue-Green Polyphenolic Pigment from the Diatom
253 *Haslea ostrearia* (Gaillon/Bory) Simonsen Responsible for the Natural Greening of Cultured
254 Oysters', *J. Agric. Food Chem.*, vol. 56, no. 15, pp. 6278–6286, Aug. 2008, doi:
255 [10.1021/jf073187n](https://doi.org/10.1021/jf073187n).
- 256 5. S. Méresse *et al.*, 'Haslea ostrearia Pigment Marenine Affects Key Actors of Neuroinflammation
257 and Decreases Cell Migration in Murine Neuroglial Cell Model', *IJMS*, vol. 24, no. 6, p. 5388,
258 Mar. 2023, doi: [10.3390/ijms24065388](https://doi.org/10.3390/ijms24065388).
- 259 6. R. Gastineau *et al.*, 'Biological Activities of Purified Marenine, the Blue Pigment Responsible for
260 the Greening of Oysters', *J. Agric. Food Chem.*, vol. 60, no. 14, pp. 3599–3605, Apr. 2012, doi:
261 [10.1021/jf205004x](https://doi.org/10.1021/jf205004x).
- 262 7. V. Turpin, J.-M. Robert, P. Gouletquer, G. Massé, and P. Rosa, 'Oyster greening by outdoor mass
263 culture of the diatom *Haslea ostrearia* Simonsen in enriched seawater: Mass production of the
264 diatom *Haslea ostrearia*', *Aquaculture Research*, vol. 32, no. 10, pp. 801–809, Oct. 2001, doi:
265 [10.1046/j.1365-2109.2001.00615.x](https://doi.org/10.1046/j.1365-2109.2001.00615.x).

- 266 8. F. Piveteau, 'Aroma of oyster *Crassostrea gigas*: effect of supplementation with the microalgae
267 *Skeletonema costatum*', Doctoral dissertation, Nantes University, 1999. Accessed: Dec. 15, 2022.
268 [Online]. Available: <https://www.theses.fr/1999NANT2017>
- 269 9. F. S. Prasetya, 'Greening Phenomenon in Bivalve by Marenine Produced from *Haslea ostrearia*
270 and its Consequences on Bivalve's Integrated Response', Doctoral dissertation, Le Mans
271 University, 2015.
- 272 10. S. La Barre and S. S. Bates, *Blue Biotechnology: Production and Use of Marine Molecules*.
273 Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2018. doi:
274 [10.1002/9783527801718](https://doi.org/10.1002/9783527801718).
- 275 11. A. Lépinay, H. Capioux, V. Turpin, F. Mondeguer, and T. Lebeau, 'Bacterial community structure
276 of the marine diatom *Haslea ostrearia*', *Algal Research*, vol. 16, pp. 418–426, Jun. 2016, doi:
277 [10.1016/j.algal.2016.04.011](https://doi.org/10.1016/j.algal.2016.04.011).
- 278 12. A. Lépinay *et al.*, 'First insight on interactions between bacteria and the marine diatom *Haslea*
279 *ostrearia*: Algal growth and metabolomic fingerprinting', *Algal Research*, vol. 31, pp. 395–405,
280 Apr. 2018, doi: [10.1016/j.algal.2018.02.023](https://doi.org/10.1016/j.algal.2018.02.023).
- 281 13. P. J. Harrison, R. E. Waters, and F. J. R. Taylor, 'A broad spectrum artificial sea water medium
282 for coastal and open ocean phytoplankton', *Journal of Phycology*, vol. 16, no. 1, pp. 28–35, Mar.
283 1980, doi: [10.1111/j.0022-3646.1980.00028.x](https://doi.org/10.1111/j.0022-3646.1980.00028.x).
- 284 14. J. F. C. De Brouwer, K. Wolfstein, and L. J. Stal, 'Physical characterization and diel dynamics of
285 different fractions of extracellular polysaccharides in an axenic culture of a benthic diatom', *Euro.*
286 *J. Phycol.*, vol. 37, no. 1, pp. 37–44, Feb. 2002, doi: [10.1017/S0967026201003419](https://doi.org/10.1017/S0967026201003419).
- 287 15. 'GitHub Source Code'. Available: https://github.com/Thrylia/16S_Haslea_ostrearia
- 288 16. 'National Center for Biotechnology Information'. Accessed: Dec. 15, 2022. [Online]. Available:
289 www.ncbi.nlm.nih.gov.

- 290 17. J. S. Johnson *et al.*, ‘Evaluation of 16S rRNA gene sequencing for species and strain-level
291 microbiome analysis’, *Nat Commun*, vol. 10, no. 1, p. 5029, Dec. 2019, doi: [10.1038/s41467-019-](https://doi.org/10.1038/s41467-019-13036-1)
292 [13036-1](https://doi.org/10.1038/s41467-019-13036-1).
- 293 18. M. Sapp, A. Wichels, and G. Gerdt, ‘Impacts of Cultivation of Marine Diatoms on the Associated
294 Bacterial Community’, *Appl Environ Microbiol*, vol. 73, no. 9, pp. 3117–3120, May 2007, doi:
295 [10.1128/AEM.02274-06](https://doi.org/10.1128/AEM.02274-06).
- 296 19. T. Ling *et al.*, ‘Analysis of bacterial community diversity within seven bait-microalgae’, *Algal*
297 *Research*, vol. 51, p. 102033, Oct. 2020, doi: [10.1016/j.algal.2020.102033](https://doi.org/10.1016/j.algal.2020.102033).
- 298 20. J. A. Kimbrel *et al.*, ‘Host selection and stochastic effects influence bacterial community assembly
299 on the microalgal phycosphere’, *Algal Research*, vol. 40, p. 101489, Jun. 2019, doi:
300 [10.1016/j.algal.2019.101489](https://doi.org/10.1016/j.algal.2019.101489).
- 301 21. Y. Ding *et al.*, ‘Response of the rearing water bacterial community to the beneficial microalga
302 *Nannochloropsis oculata* cocultured with Pacific white shrimp (*Litopenaeus vannamei*)’,
303 *Aquaculture*, vol. 542, p. 736895, Sep. 2021, doi: [10.1016/j.aquaculture.2021.736895](https://doi.org/10.1016/j.aquaculture.2021.736895).
- 304 22. J. Lian, P. Schimmel, S. Sanchez-Garcia, R. H. Wijffels, H. Smidt, and D. Sipkema, ‘Different co-
305 occurring bacteria enhance or decrease the growth of the microalga *Nannochloropsis* sp.
306 CCAP211/78’, *Microb. Biotechnol.*, vol. 14, no. 3, pp. 1159–1170, May 2021, doi: [10.1111/1751-](https://doi.org/10.1111/1751-7915.13784)
307 [7915.13784](https://doi.org/10.1111/1751-7915.13784).

L'article détaille les résultats obtenus pour 4 échantillons, dont 2 provenant chacun d'une culture réalisée selon un protocole classique et 2 autres selon un protocole légèrement modifié (2 x 2 échantillons pour un protocole donné, 1 échantillon = 1 culture). En raison du nombre insuffisant de réplicats, les données obtenues n'ont pas permis d'aboutir à une conclusion définitive sur l'effet des antibiotiques sur la population bactérienne des cultures. En effet, il aurait été nécessaire d'avoir un troisième réplicat biologique, c'est-à-dire une troisième culture pour chacun des 2 protocoles, pour réaliser des tests statistiques et mesurer les différences dans les compositions bactériennes. Mais comme *H. ostrearia* est difficile à cultiver, seules 4 cultures ont pu être produites et utilisées pour émettre l'hypothèse qu'un traitement antibiotique influence les concentrations des bactéries et les stabilise. Lorsque ce dernier n'est pas inclus dans le protocole de culture, on observe une variation de ces dernières. Un grand nombre de bactéries a également été retrouvé, réparties dans au minimum 35 ordres différents, ce qui a augmenté nos soupçons précédents sur la grande diversité bactérienne présente dans les cultures de *H. ostrearia*. Cependant, toutes ces bactéries ont été associées à un taxon de bactéries marines, ce qui élimine l'idée d'une possible contamination des cultures en laboratoire.

Il a été également noté que les taxons bactériens retrouvés avec et sans traitement antibiotique ne semblent pas différer énormément. En regardant de plus près le tableau des "Supplementary data" de l'article, il peut être remarqué que, par rapport à la diversité observée, peu de bactéries n'ont été retrouvées que chez les cultures traitées (6 spécifiques à T) ou que chez celles non traitées (5 spécifiques à N). Les antibiotiques utilisés ne semblent pas avoir un effet important sur la diversité des bactéries observées lors de cette étude, mais il serait nécessaire d'étudier un plus grand nombre de données pour proposer une réelle conclusion à cette observation.

Enfin, cette étude a permis de mettre en avant l'attention particulière à apporter lors de la sélection de la région 16S à séquencer pour l'identification bactérienne (Figure 2-1). En effet, la région V1V3, très fréquemment utilisée, semblait avoir des lacunes pour détecter des bactéries telles que celles appartenant à la classe des *Gammaproteobacteria*, ce qui n'a pas

été le cas de la région V3V4. Si nous n'avions pas vérifié les aptitudes des régions sélectionnées V1V3 et V3V4 à les détecter, nous aurions pu passer à côté d'une classe très présente lors de la culture en laboratoire de microalgues (Sapp et al 2007) telles que *H. ostrearia*.

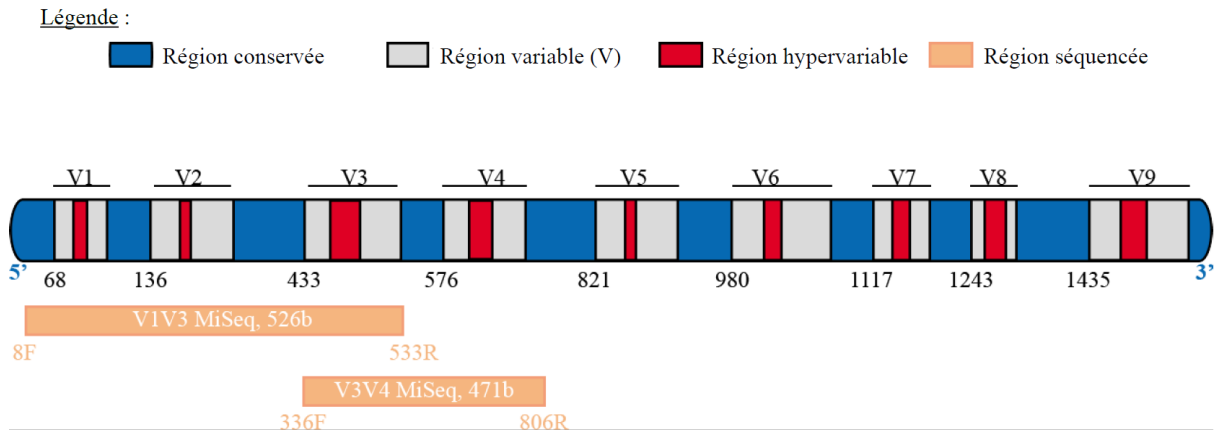


Figure 2-1 : Représentation schématique des régions variables et hypervariables de l'ARN ribosomique 16S et des régions séquencées pour l'étude

Les amorces utilisées pour le séquençage des régions V1V3 et V3V4 sont indiquées sur la Figure et correspondent à celles données par Sun et al (2014) et Munyaka et al (2015), respectivement.

3. Conclusion du chapitre

De nombreuses bactéries marines ont été identifiées grâce au séquençage de la région V3V4 du gène ARNr 16S, dont certaines avaient déjà été associées à une amélioration de la croissance d'autres espèces de microalgues (Lian et al 2021). Ces dernières ne semblent pas provenir d'une contamination extérieure des cultures, mais de l'environnement naturel de *H. ostrearia*. Il est donc fortement soupçonné que ces dernières offraient un avantage à la microalgue, et lui permettent de survivre dans certains milieux. Ceci n'est pas sans rappeler la tendance observée en laboratoire des cultures *H. ostrearia* à ne plus proliférer lors d'un traitement antibiotique trop puissant ou trop long. Ceci suggère que la survie de *H. ostrearia* dépendrait de la présence de certaines bactéries, appelées "bactéries nécessaires" pour la suite. Ce qui justifie, de nouveau, la nécessité de considérer la présence de ces dernières dans les protocoles de culture en laboratoire.

Ainsi, les variations observées dans les concentrations bactériennes entre un protocole avec et sans traitement antibiotique semblent démontrer des résultats prometteurs pour contrôler cette communauté bactérienne. Cependant, le traitement utilisé ici ne donne pas l'impression de réduire suffisamment la présence des bactéries ou leur diversité pour atteindre notre objectif de purification des cultures. En effet, bien que leur concentration semble se stabiliser et offrir une possibilité d'obtenir la même population bactérienne à chaque nouvelle culture (reproductibilité des cultures), leur diversité n'en était pas moindre comparée à celle obtenue dans des cultures dites "classiques", sans traitement. Il a donc été proposé d'étudier les génomes de ces bactéries précédemment obtenus afin d'en apprendre plus sur celles-ci, et notamment comment les éliminer.

Toutefois, il y a une forte probabilité que la communauté bactérienne présente dans nos cultures ait évolué avec le temps (Lépinay et al 2016). En effet, les microalgues à l'origine de notre culture ont été isolées en 2018 et les travaux de séquençage n'ont eu lieu qu'à partir de 2020, la proportion occupée par chacune des bactéries en a très probablement été impactée. Des bactéries supposément rares dans nos cultures pourraient être très présentes

à l'état naturel, ou inversement. Or l'intérêt ici n'était pas de retrouver les concentrations des bactéries présentes à l'état naturel, mais bien de les identifier dans le cadre des cultures pour améliorer le protocole et pouvoir récupérer le génome de *H. ostrearia* par la suite.

Chapitre 3 : Etude des gènes de résistance aux antibiotiques chez les bactéries des cultures de *H. ostrearia*

1. Contexte des travaux de recherche

Suite aux précédents tests, il est apparu que les cultures de *H. ostrearia* étaient associées à une grande diversité bactérienne. Ces bactéries marines, bien que contaminantes de notre point de vue pour l'identification du génome de *H. ostrearia*, semblent appartenir, pour certaines, à des groupes taxonomiques précédemment identifiés comme utiles à la croissance des microalgues. De plus, selon les observations faites lors de tests de décontamination totale des cultures, *H. ostrearia* semblait ne pas survivre sans elles. L'hypothèse de l'existence d'un consortium bactérien nécessaire à la survie de *H. ostrearia* a donc été émise. Devant cette supposition, la nécessité de pouvoir éliminer des bactéries précises des cultures afin de pouvoir identifier lesquelles seraient susceptibles d'avoir un rôle important dans la vie de la microalgue est apparue. Si une bactérie X est absente des cultures et que *H. ostrearia* meurt, alors cette dernière serait, hypothétiquement, nécessaire. Au contraire, si la microalgue survit alors la bactérie X peut être éliminée de nos cultures. Ceci permettrait notamment, à terme, de réduire la diversité bactérienne et de faciliter l'accès au génome de *H. ostrearia*, puisque l'hétéroclisme des bactéries présentes dans les cultures n'a pas été réduit par l'utilisation d'antibiotiques à large spectre dans nos précédentes analyses.

Mais, pour pouvoir cibler des bactéries données, il faut pouvoir employer des moyens spécifiques à ces dernières. L'utilisation d'antibiotiques spécifiques aux taxons bactériens identifiés est apparue comme une piste à privilégier. Pour cela, il a été proposé de lister les gènes conférant une résistance ou une insensibilité aux antibiotiques (ARGs) et aux substances antimicrobiennes (AMRs) pour chaque famille bactérienne. Ces gènes ont été retrouvés dans les génomes bactériens précédemment assemblés, et ce grâce à deux approches différentes : Prokka (Seemann 2014) et RGI vs la base de données CARD (Alcock

et al 2019, 2023). L'entièreté du protocole d'analyse et des résultats sont disponibles dans l'article suivant.

2. Article sur la recherche de AMRs et ARGs

[Article prochainement soumis à PCI Microbiol]
[Les *Supplementary data* de l'article sont disponibles dans les Annexes]

1 **Identification of antibiotic-resistance genes in order to reduce the bacterial**
2 **content of *Haslea ostrearia* cultures**

3

4 **Authors**

5 Aurelie Peticca^{a*}, Benoit Chenais^a, Nathalie Casse^{a §} and Francois Sabot^{b §}

6 ^a BiOSSE (Biology of Organisms: Stress, Health, Environment), UFR Sciences et Techniques, Le Mans
7 Universite, F-72085 Le Mans, France

8 ^b DIADE, University of Montpellier, CIRAD, IRD, F-34394 Montpellier, France

9 *corresponding author: aurelie.peticca@univ-lemans.fr

10 §Authors with equal contribution

11

12

13 **Abstract**

14 *Haslea ostrearia* is a microalga complex to cultivate axenically in the laboratory, requiring a mandatory
15 bacterial community, making its genome analysis difficult and contributing to a lack of understanding
16 on it. In previous laboratory tests, the addition of broad-spectrum antibiotics to the usual culture protocol
17 has been unsuccessfully used to reduce the number of bacteria in cultures. Here, we investigated the
18 resistance and sensitivity to antibiotics of the most commonly present bacteria in the samples through
19 the presence of resistance genes. This allowed us to draw up a non-exhaustive list of antibiotics to be
20 tested in priority and those to be avoided, as well as a potential explanation on the ineffectiveness of the
21 currently used antibiotics. Finally, a brief analysis of the evolution of the genes conferring resistance,
22 highly present in the data, was proposed, allowing us to estimate whether these resistances were acquired
23 before or during microalgae culture.

24

25 **Keywords**

26 diatoms, bacteria, antibiotic resistance, *Haslea ostrearia*, bacterial community, microalgae

27

28 **1. Introduction**

29 *Haslea (H.) ostrearia* is a microalgae of the diatom family (*Bacillariophyta*), well known for its blue
30 pigment called the marennine (Simonsen 1974, Gastineau et al 2014). It is very common in the oyster
31 beds of the French Atlantic coast, where it gives the water a greenish appearance (Gastineau et al 2012,
32 Turpin et al 2001). The main interests in this microalgae are the potential uses of its pigment in the
33 medical field for its antioxidant effect, which could be beneficial to humans (Pouvreau et al 2008), and
34 the changes it brings to oysters, as these changes increase the market value of these bivalves (La Barre
35 and Bates 2018).

36 However, *H. ostrearia* is still a poorly known species. For instance, it is still difficult to say how and
37 under which circumstances the marennine pigment is produced. The *H. ostrearia* genome has not yet
38 been partially assembled or even properly sequenced. Indeed, due to its complex ecology, *H. ostrearia*
39 seems to be difficult to grow in an axenic medium, as it seems to require a rather complex bacterial
40 community (Lepinay et al 2018, Peticca et al 2023b), which contaminates any attempt so far to obtain
41 *H. ostrearia* DNA. Therefore, various culture protocols including the presence of broad-spectrum
42 antibiotics have been proposed, but many bacteria are still found in treated samples.

43 Thus, this study proposes to search for antimicrobial resistance genes (AMRs), and more specifically
44 for antibiotics resistance genes (ARGs) in the microbial genomes associated with a representative *H.*
45 *ostrearia* culture, in order to improve culture protocols by proposing efficient antibiotics. This could
46 allow *H. ostrearia* to be grown with a reduced number of different bacterial species, making its genome
47 easier to study, and then offering a better understanding of *H. ostrearia* physiology.

48

49 For this purpose, *H. ostrearia* was grown in the laboratory using a standard protocol (Peticca et al 2023a,
50 Peticca et al 2023b). All DNA present in the samples was sequenced using Illumina and
51 PacificBiosciences (PB) technologies. The metagenome was assembled and the bacterial genomes were
52 identified through an *in silico* 16S rDNA search. AMRs and ARGs were then annotated in the best
53 assembled genomes using Prokka (Seemann 2014) and RGI software (Alcock et al 2020) against the
54 CARD database (The Comprehensive Antibiotic Resistance Database, Alcock et al 2023). Antibiotic
55 resistances found in at least 2/3 of the bacterial genomes, and those very rarely found were scored to
56 provide a non-exhaustive list of the potentially least and most effective antibiotics. Finally, the evolution
57 of genes associated with resistance to antibiotics, present in the majority of bacterial genomes, was
58 studied in order to verify whether these gene modifications were recent or ancient, and therefore
59 preceded the cultivation of the microalgae or not.

60

61 **2. Material and Methods**

62 All the scripts are available on GitHub (https://github.com/Thrylia/Haslea_AMR_research, GNU
63 General Public License v3.0). DNA sequences and the best assembled genomes sequences (Busco score
64 higher than 50%) were deposited in NCBI database under the BioProject PRJNA843895. The
65 BioSample accession numbers for each of those genomes are included in the Additional Table 1.

66 **2.1 From the culture of *Haslea ostrearia* to the extraction of its DNA**

67 Registered as NCC 532 in the Nantes Culture Collection (Vona Meleder, vona.meleder@univ-nantes.fr
68 ; Nantes, France), the isolate selected for this study was issued from a 2018 oyster bed in Bouin (France,
69 46°57'12.4"N, 2°02'46.1"W). Sequencing was performed by Genotoul (Toulouse, France) using the
70 PacBio continuous long reads (PB CLR SEQUEL2) and Illumina MiSeq platforms (TrueSeqv3, 250pe).
71 Respectively, 85Gb and 5.2Gb total read length were generated during the process (Peticca et al 2023a).

72 **2.2 DNA treatment, assembly and analyzes**

73 The Illumina read quality was checked with FastQC 0.11.8 and formatted using MultiQC 1.7 (S.
74 Andrews et al 2010, Ewels et al 2016). Trimmomatic 0.39 was used (Bolger et al. 2014) to remove the
75 remaining adapter sequences, low quality or short reads (<Q30, <75b) from Illumina set. The PacBio
76 reads were assembled with Flye 2.9 in meta mode and with the -g 100m option (Kolmogorov et al.
77 2020). The obtained contigs were afterwards corrected with the filtered Illumina reads and three loops
78 of racon 1.4.20 and bwa-mem2 2.2.1 (Vaser et al 2017, Vasimuddin et al. 2019). The average depth of
79 each sequence and their circularity were recovered from the files produced by Flye. To identify the
80 bacteria in the reconstructed metagenome, an *in silico* PCR was performed to get the positions of
81 complete 16S rDNA genes using universal primers 27F-AGAGTTTGATCMTGGCTCAG and 1492R-
82 TACGGYTACCTTGTTACGACTT (Heuer et al 1997), with Blastn 2.12.0 and the option -word_size
83 6 (Altschul et al 1990). The 16S rDNA sequences were then extracted with samtools faidx 1.15.1
84 (Danecek et al. 2021), and assigned to a bacterial taxon using SILVA database 138 Ref NR 99 and
85 Qiime2 2021.2.0 with the feature classifier classify-consensus-blast (Quast et al 2012, Bolyen et al
86 2019). The bacterial genomes were then annotated by Prokka 1.14.6 (Seeman 2014), and their
87 completeness was calculated using BUSCO 5.4.3 (Manni et al 2021). Only genomes with a BUSCO
88 percentage (Busco score) greater to 50% and a depth greater to 20X were screened for AMRs. A search
89 for the “resistance” keyword was conducted on the gff files created by Prokka. A specific ARGs
90 annotation using RGI 6.0.1 software (Alcock et al 2020) against the CARD 3.2.5 database (The
91 Comprehensive Antibiotic Resistance Database, Alcock et al 2023) was performed. Only ARGs with an
92 identity percentage of 80% or more and a coverage percentage of at least 70% were retained. Finally, in
93 order to verify that the ORFs of the RGI annotated ARGs matched the Prokka annotation, only those
94 ARGs that matched a Prokka ORF with a minimum of 70% coverage were retained for further
95 processing. The latter filter was performed using bedtools intersect 2.30.0 (Quinlan and Hall 2010). The
96 search for AMRs was made directly on these genomes using the Prokka annotation as previous analyses
97 already demonstrated the effectiveness of Flye for the study of resistant genes (Brown et al 2021,
98 Kolmogorov et al 2020). Prokka was also proved to be robust in its annotation compared to other
99 pipelines actually available (Karimi et al 2021).

100 **2.3 Visualisation**

101 The plot figures were all made using R 4.2.3 and ggplot2 3.4.0, except for the Sankey plot which was
102 created with network3D 0.4 and htmlwidgets 1.6.2. The scatter plot used the data generated by Flye
103 (mean depth, genome circularity), by Busco (Busco score) and by Prokka (AMRs found) to make an
104 overall profile of the genome assembled. For the Presence-Absence matrix, only the AMRs and ARGs
105 annotated by RGI and Prokka were exploited to highlight the high presence of antimicrobial and
106 antibiotic resistance genes in the genomes. The Sankey plot required sorting the data for ARMs and
107 ARGs into specific categories of mechanisms (Munita and Arias 2016) before its creation to enhance
108 the resistances with the most associated genes in the genomes and associated mechanisms. Phylogenetic
109 trees were constructed using the default protocol of phylogeny.fr (feb 2023, Dereeper et al 2008) in
110 order to verify that the ARGs present in multiple copies in the genomes are not due to an assembly error
111 (Fig.4, Additional Fig. 1). The ARGs sequences were extracted with samtools faidx using the positions
112 found by Prokka, and the genes tagged as putative were discarded from the results.

113

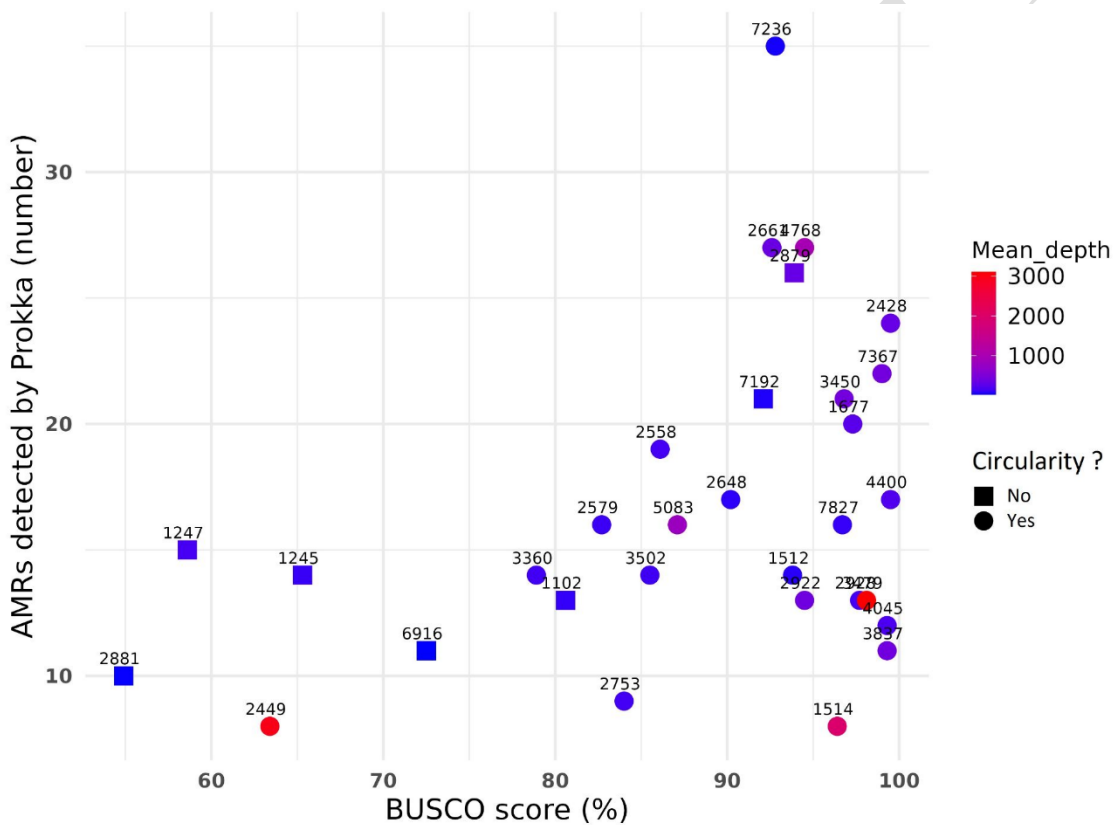
114 **3. Results and Discussion**

115 **3.1 Bacterial genomes**

116 The metagenomic assembly of the PacBio reads and its polishing with the Illumina reads resulted in a
117 dataset of 5,835 contigs. These included the already identified chloroplastic and mitochondrial genomes
118 of *H. ostrearia* (Peticca et al 2023a) and 77 contigs with a complete 16S sequence found in the *in-silico*
119 PCR step. From these, only the 31 ones with a Busco score greater than 50% (relative to the bacterial
120 class to whom they belong), and a mean depth equal or greater than 20X were retained for further
121 analyses (Additional Table 1). These filters removed a possible lack of results for the resistance genes
122 research due to a missing extended sequence, as well as the possibility to overlook on poorly present
123 bacteria in the culture. Indeed, the main aim of this study was to identify suitable antibiotics to purify
124 the cultures of as many bacteria as possible, *i.e.* of the most prevalent ones.

125 Twenty-four of these 31 genomes were circular, but the length and the high Busco score of the remaining
 126 7 indicated that they were missing only small sequences (Fig. 1). Circular ones had a Busco score above
 127 60%, the majority (17) being between 90% and 100%. Of all these genomes, two stood out by the taxon
 128 that has been associated with them: #7236 (Busco score > 90% and circular) and #6916 genomes (Busco
 129 score > 70% but not circular) from KI89A clade (Cho and Giovannoni 2004) and NS11-12 marine group
 130 (Meziti et al. 2015), respectively. These taxa did not have an associated genome yet, probably because
 131 of their relatively recent discovery and the difficulty of growing them in the laboratory.

132 **3.2 Identification of antimicrobial resistance by Prokka and RGI**



133
 134 **Figure 1: Scatter plot of the AMRs detected by Prokka for the bacterial genomes associated to a**
 135 **Busco score greater than 50%**

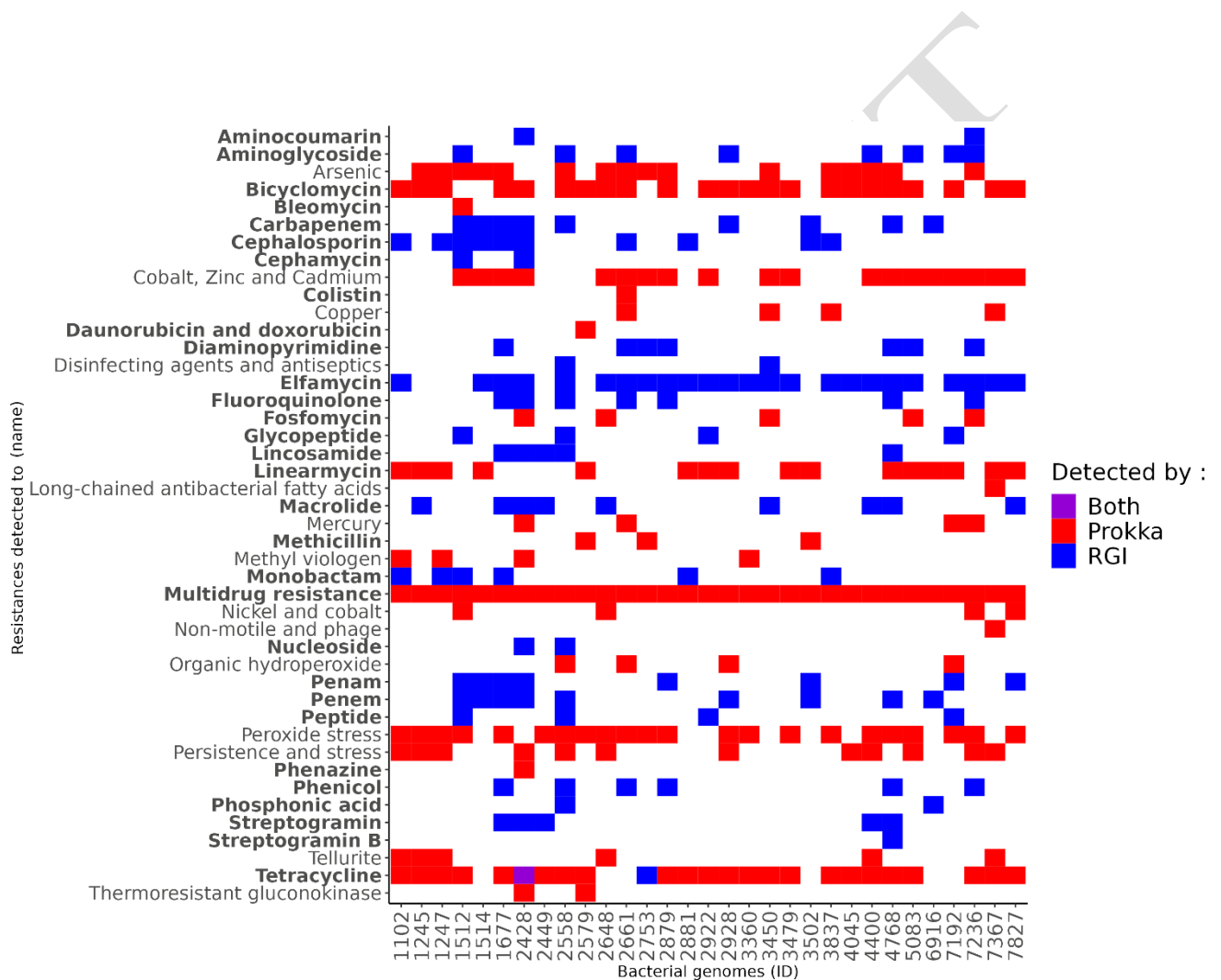
136 The mean depth, calculated by Flye during the assembly step, is represented by the color gradient. AMRs were
 137 detected by Prokka and searched using the word “resistance” as a marker. The identification number (ID) of each
 138 bacterial genome in the assembly is written in black above its associated point, with the shape indicating where
 139 the genome is circular (circle) or non-circular (square).

140

141 As expected, the number of AMRs tended to increase with the Busco score (Fig. 1), because if the
142 assembly is complete or nearly complete, fewer AMRs will be missed by Prokka. Only the #7236
143 genome exceeded the 30 genes found, the average being 15 with a minimum of 3 and a second-to-last
144 maximum of 27 genes.

145

146



147

148 **Figure 2: Presence-absence matrix of resistance detected by Prokka or RGI or both in the**
149 **different genomes**

150 Resistances to a certain type of molecule, components, mechanisms or stress identified by Prokka are shown in
151 red, those identified by the RGI pipeline used against the CARD database in blue, and those detected by both are

152 marked in purple. Resistances to antibiotic molecules or related processes are shown in bold. The bacterial genome
153 identifiers (IDs) used are the genome identifiers in the metagenomic assembly.

154

155

156

157

158

159

160

161

162

163

164

165

166

167

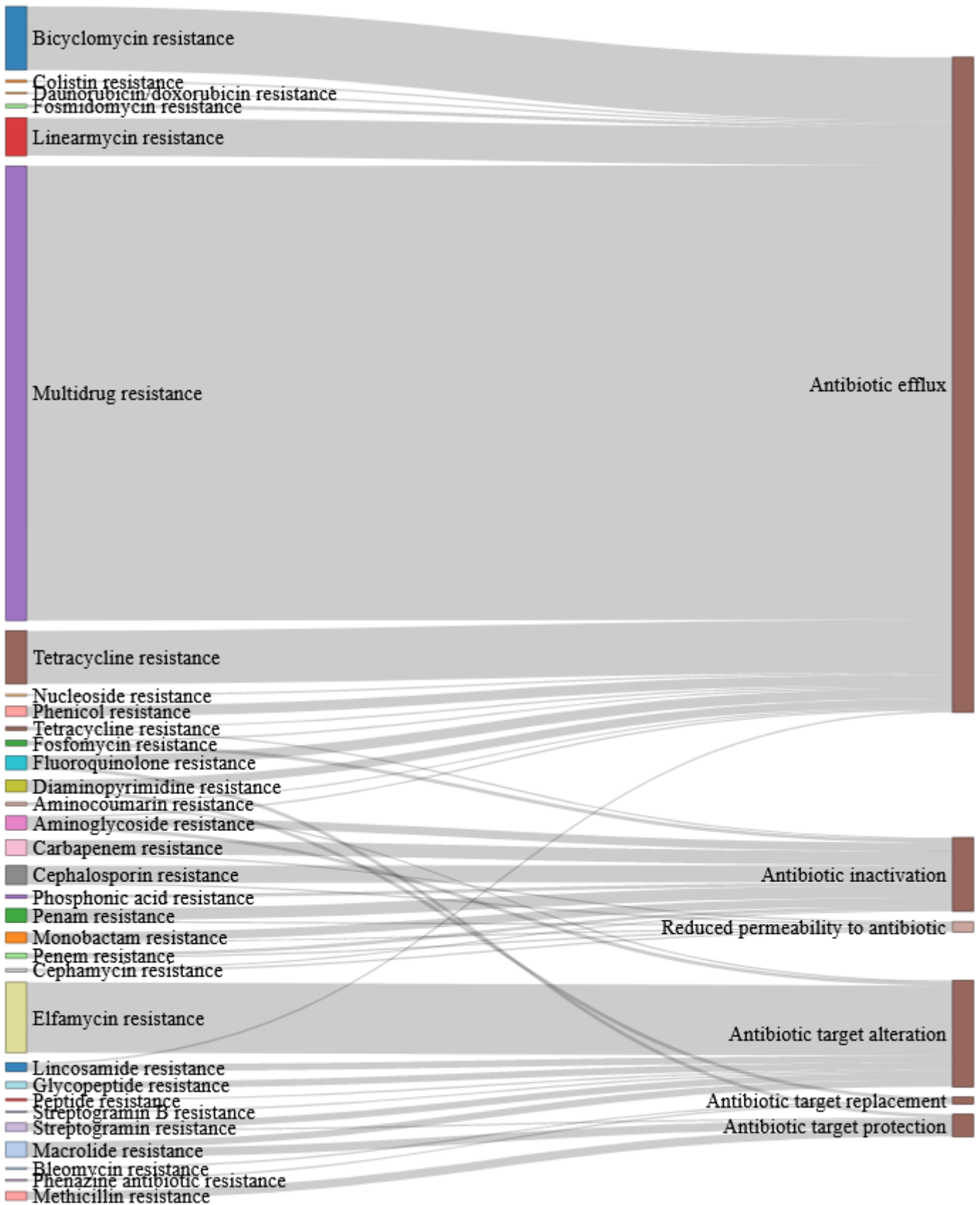
168

169

170

171

PREPRINT



173 **Figure 3: Sankey of the types of mechanisms impacted for the acquisition of antibiotic resistance**

174 **identified by Prokka or RGI or both**

175 Each gene linked to antibiotic resistance and identified by Prokka or RGI or both was looked for in the genomes
176 and categorised according to the mechanism impacted by its modification. The categories were retrieved from the
177 RGI pipeline against the CARD database and echo those used by other studies (Munita and Arias 2016). The total
178 number of genes found is the flow thickness.

179

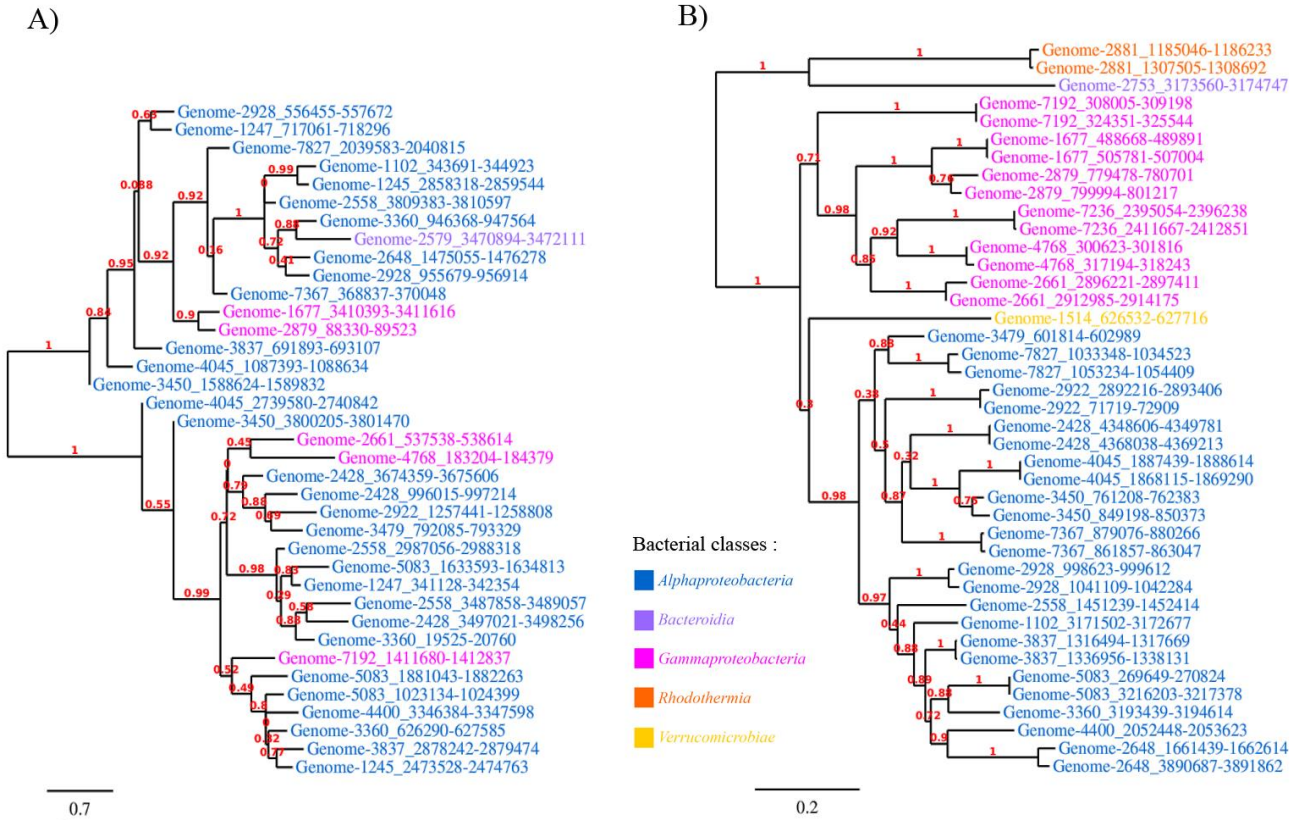
180 To further investigate ARGs specific results, the RGI pipeline was run against the CARD database to
181 search for genes that confer resistance to currently used antibiotics, and the resistances found added to
182 those obtained by Prokka (Fig. 2). Only the #2428 genome was found to be resistant (Tetracycline
183 antibiotic) by both programs. It could be explained by the fact that Prokka and RGI do not operate in
184 the same way. The first one annotates a prokaryotic genome in a general way and is based on Prodigal
185 (Hyatt et al 2010) for coding sequence search, followed by a Blast+ (Camacho et al 2009) against the
186 SWISS-prot database (Bairoch and Apweiler 1997) to proceed to a protein annotation (Seeman 2014).
187 Although it also uses Prodigal for the coding sequence detection, RGI aligns these using DIAMOND
188 (Buchfink et al 2021) against the CARD database (Alcock et al 2020) . As the databases are different, it
189 is normal to find different results. The filters associated with each of these pipelines also differ. In this
190 study, the two approaches were used with the idea of complementing each other rather than being
191 compared.

192 Overall, at least 48 molecules/mechanisms/components were potentially inefficient against one or more
193 of the bacteria in the culture (Fig. 2). Although multidrug resistance appeared for all the 31 genomes
194 analyzed, these multidrug efflux pumps are well-known antibiotic resistances proteins found in all
195 microorganisms, and are thus not very informative and therefore were not further considered in this
196 study (Fig. 3; Blanco et al 2016). Besides, 30 antibiotics were detected, with the most represented
197 resistances being to elfamycin (24 genomes upon 31), bicyclomycin (23/31) and tetracycline (22/31).
198 Thus, it is possible to advise against the use of the other 3 antibiotics listed earlier on the basis of these
199 results. In addition, more than 2/3 of the genomes have accumulative resistance to elfamycin,
200 bicyclomycin or tetracycline or both antibiotics. Their efficiency in reducing the number of bacteria in
201 cultures is therefore questionable. This does not seem to be the case with antibiotics for which few

202 genomes are resistant, as bleomycin (1 genome out of 31), phenazine (1/31), daunorubicin and
203 doxorubicin (1/31), and colistin (1/31). In addition, it is worth noting that there was evidence of
204 resistance to streptogramin for 5 genomes: this class of antibiotics may therefore not be as effective as
205 the others mentioned above, and should not be recommended for further analysis.

206 **3.3 Identification of several new resistances compared with previously available data**

207 These results could also explain why the Penicillin-Streptomycin-Amphotericin β suspension currently
208 used in the standard culture protocol was not as efficient as expected. Beyond the effect that the
209 multidrug resistance could have on treatment with its efflux pump mechanism, as previously stated
210 (Blanco et al 2016), the present study showed specific resistance to each of the antibiotics (Fig. 2, Fig.
211 3). Firstly, penicillin belongs to the group of β -lactam antibiotics, and it has already been shown that
212 resistances to other β -lactam antibiotics were usually associated with some degree of non-susceptibility
213 to penicillin (McGee et al 2015). In this study, resistances to the β -lactam antibiotics penam and penem
214 (penicillin derivatives), cephalosporin, cephamycin, monobactam and carbapenem were found across
215 18 different genomes (Fig. 2), the mechanism associated with all of them being their inactivation (Fig.
216 3). A second mechanism was also observed for penam, penem, cephamycin and monobactam, with a
217 reduced permeability to them. Further analyses also revealed the presence of modified *pbp* genes in the
218 annotations of 16 genomes, which could act as an indicator of penicillin resistance (McGee et al 2015).
219 For streptomycin, a broad-spectrum aminoglycoside antibiotic (PubChem, Streptomycin Compound
220 Summary) resistance to such class was found in 8 genomes (Fig. 2). This resistance would mainly be
221 due to an inactivation of the antibiotics, and to a lesser extent to a reduced permeability to them and an
222 efflux pump (Fig. 3). Lastly, amphotericin β is an antifungal antibiotic from the macrolide class
223 (PubChem, Amphotericin B Compound Summary), to which 9 genomes were annotated as resistant to,
224 with alteration and the protection of the antibiotic target (Fig. 2, 3).



22
 226 **Figure 4: Phylogenetic trees of the modified the *bcr* gene (A) and the *EF-Tu* gene (B) from the**
 227 **bacterial genomes identified in *H. ostrearia* culture**

228 A) Phylogenetic tree corresponding to the gene *bcr*, which confers resistance to bicyclomycin antibiotic. B)
 229 Phylogenetic tree for a modified *EF-Tu* gene that confers resistance to elfamycin antibiotic. Each sequence is in
 230 the format “Genome-*{ID}*_*{start}*_*{stop}*”, which corresponds to the genome identified in the metagenomic
 231 assembly, and the start and stop position of the gene of interest. The leaves color indicates the class to which the
 232 bacterial genome from which each gene comes belongs. The phylogenetic trees were made by phylogeny.fr.

233
 234 In order to verify that the resistances found to be the most abundant in the genomes have not been
 235 acquired recently, phylogenetic trees were estimated from the modified *EF-Tu* and *bcr* genes sequence
 236 (Fig. 4), conferring resistance to elfamycin and bicyclomycin, respectively. For the *bcr* gene
 237 phylogenetic tree, there is no correlation between the taxon associated to the genome and its localisation
 238 on the tree (Fig. 4 - A), but it turned out that this gene was present in two copies in almost half of the

239 genomes: #1245, #1247, #2428, #2558, #2928, #3360, #3450, #3837, #4045 and #5083. This is probably
240 the result of an ancient duplication of this gene, as shown by the splitting of the tree at its base, and the
241 placement of each of the orthologous copies on either side (duplication in a common ancestor); the
242 branches length also supports this explanation: representing a genetic change of 0.7, the duplication of
243 this gene was followed by numerous mutations, representing a long time of evolution. A phylogenetic
244 tree was also made for tetracycline and the modified *tetA* gene, and gave the same kind of results as for
245 the *bcr* tree (Additional Fig. 1). These observations were not the same for the *EF-Tu* tree (Fig. 4B),
246 where the leaves of the tree appeared to be grouped by taxonomy, formed of closely related nodes and
247 very short branches between copies of the same genome. This pattern could be a sign of poor genome
248 assembly and of an artificial duplication by the assembler of a region harbouring the gene. However,
249 this hypothesis was disproved by comparing the ends of each copy of the gene in the same genome: the
250 1,000 bases before and after the gene were never similar between each copy (Additional Table 2). The
251 explanation could therefore lie in the evolution of the *EF-Tu* gene (also known as *Tu*), which is highly
252 conserved in genomes and is even used for species discrimination (Harvey et al. 2019). This strong
253 discriminatory potential is reflected in the grouping of our genes into bacterial class clusters: a potential
254 duplication may have occurred eons ago, and the selection pressure on the gene sequences acted then to
255 conserve similar sequences in the same genome. The two copies of the gene differ from each other by
256 less than 1.4% of their nucleotide sequence, which could be the reason why the branches on the tree
257 were extremely short and the copies of the gene grouped by genomes.

258 According to these analyses, it seemed that these antibiotic resistances were ancient. The lengths of the
259 branches are important, indicating a large number of mutations in the genes studied. Analysis shows that
260 there has been no recent horizontal transfer and therefore no acquisition due to laboratory culture of *H.*
261 *ostrearia*. This means that the resistances associated with *H. ostrearia* cultures could also be observed
262 in their natural environment. It is therefore very possible, as mentioned above, that these antibiotics will
263 never have any effect on these bacteria, as these resistances are not uncommon to find (Selvarajan et al
264 2023, Olsthoorn-Tieleman et al 2007, Wyres et al 2013), even for the resistances to currently used
265 antibiotics in the culture protocol. The first resistance to penicillin in a bacteria was first noticed in 1942

266 (Dodds 2017) and a worldwide emergence was noted in the 1980s. This resulted in the massive use of
267 macrolide antibiotics, among others. To such an extent that, in the 1990s, macrolide resistance was even
268 found more frequent than the latter in most parts of the world (Farrel et al 2002). In recent years, an
269 emergence of the resistance to aminoglycosides was also revealed (Doi et al 2016). What is more
270 worrying is the presence of resistance to so-called drug of last resort (DoLR): the results showed
271 potential insensitivity to colistin in the 2661 genome and to fluoroquinolone in 7 different genomes. The
272 bacteria identified here that possess these resistances do not appear to be pathogenic, and therefore do
273 not represent a problem on their own. The concern would be if these bacteria encountered
274 dangerous/deadly bacteria and passed on these resistances, especially since this isolate of *H. ostrearia*,
275 in which these bacteria were found, came from an oyster farm. Oysters that feed on this microalga could
276 transport these bacteria out of the water to the human environment, for instance.

277

278 **4. Present & future prospects**

279 Experimental analyses should be carried out on the effect of the antibiotics on bacteria associated with
280 *H. ostrearia* cultures, few of which have been detected as being resistant to these antibiotics. If the
281 results are as expected, *i.e.* a reduction in the number of bacterial species without killing the microalgae,
282 a new protocol could be proposed to the scientific community. If not, further research may be needed,
283 particularly into the bacteria that may play an important role in *H. ostrearia* survival, by testing the
284 various antibiotics identified here to eliminate specific bacteria and to observe the effect of their absence
285 on the survival of this microalgae. In this way, it would be possible to reduce the amount of bacteria in
286 the cultures by conserving only those that are essential, and provide an easier access to the *H. ostrearia*
287 genome. Analyses of genes present in the bacteria identified as indispensable for *H. ostrearia*, as well
288 as its own genome, may also inform us as to why they are necessary, and also on other mechanisms such
289 as the production of marennine, which is a compound produced solely by the microalgae, by the bacteria
290 or both?

291 Given these results, it would also be helpful to test if this bacterial community, and its resistance, is the
292 same in other *H. ostrearia* cultures and in natural environment, again with the idea of creating a new
293 and more suitable culture protocol, but this time for all cultures of the different isolates or strain.
294 However, these tests could also be used to compare whether the bacteria required for the survival of the
295 microalgae varies according to its location or its strain or both.

296

297 **Author Contribution**

298 Conceptualization: AP, FS; Formal analysis: AP, FS; Funding acquisition: NC; Investigation: AP, FS;
299 Methodology: AP, FS; Project administration: NC, FS; Supervision: NC, BC, FS; Visualization: AP,
300 FS; Roles/Writing - original draft: AP; Writing - review & editing: NC, BC, FS.

301

302 **Ethical approval**

303 None required.

304

305 **Disclosure statement**

306 The authors reported no potential conflict of interest.

307

308 **Funding**

309 This work was supported by funding from Région Pays de la Loire (LANCOM project) and through A.
310 Peticca's Ph.D. grant from Région Pays de la Loire and Le Mans metropole.

311

312 **Acknowledgements**

313 The authors are indebted to Genotoul, which produced all the sequencing data. The authors also
314 acknowledge Sebastien Santini (CNRS/AMU IGS UMR7256) and the PACA Bioinfo platform for the
315 availability and management of the phylogeny.fr website, as well as Dr. H el ene Gateau and Dr. Mostefa
316 Fodil for the DNA extraction and the culture protocol of *H. ostrearia*.

317

318 **Data availability statement**

319 All the scripts are available on GitHub (https://github.com/Thrylia/Haslea_AMR_research).

320 The genome sequences that supports the findings of this study are openly available in GenBank of NCBI
321 at <https://www.ncbi.nlm.nih.gov/> under the accession given in the Additional Table 1. The associated
322 BioProject number are PRJNA843895. The PacBio and Illumina raw sequencing used are openly
323 available under the SRA number SRX15503190 and SRX15503189, respectively.

324

325 **References**

326 Alcock, B.P. *et al.* (2019) ‘CARD 2020: antibiotic resistome surveillance with the comprehensive
327 antibiotic resistance database’, *Nucleic Acids Research*, p. gkz935. Available at:
328 <https://doi.org/10.1093/nar/gkz935>.

329 Alcock, B.P. *et al.* (2023) ‘CARD 2023: expanded curation, support for machine learning, and resistome
330 prediction at the Comprehensive Antibiotic Resistance Database’, *Nucleic Acids Research*, 51(D1),
331 pp. D690–D699. Available at: <https://doi.org/10.1093/nar/gkac920>.

332 Altschul, S.F. *et al.* (1990) ‘Basic local alignment search tool’, *Journal of Molecular Biology*, 215(3),
333 pp. 403–410. Available at: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).

334 Andrews, S. (2010) ‘FastQC: A Quality Control Tool for High Throughput Sequence Data’. Available
335 at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

336 Bairoch, A. and Apweiler, R. (1997) 'The SWISS-PROT protein sequence data bank and its supplement
337 TrEMBL', *Nucleic Acids Research*, 25(1), pp. 31–36. Available at:
338 <https://doi.org/10.1093/nar/25.1.31>.

339 Blanco, P. *et al.* (2016) 'Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance
340 Determinants', *Microorganisms*, 4(1), p. 14. Available at:
341 <https://doi.org/10.3390/microorganisms4010014>.

342 Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence
343 data', *Bioinformatics*, 30(15), pp. 2114–2120. Available at:
344 <https://doi.org/10.1093/bioinformatics/btu170>.

345 Bolyen, E. *et al.* (2019) 'Reproducible, interactive, scalable and extensible microbiome data science
346 using QIIME 2', *Nature Biotechnology*, 37(8), pp. 852–857. Available at:
347 <https://doi.org/10.1038/s41587-019-0209-9>.

348 Brown, C.L. *et al.* (2021) 'Critical evaluation of short, long, and hybrid assembly for contextual analysis
349 of antibiotic resistance genes in complex environmental metagenomes', *Scientific Reports*, 11(1), p.
350 3753. Available at: <https://doi.org/10.1038/s41598-021-83081-8>.

351 Buchfink, B., Reuter, K. and Drost, H.-G. (2021) 'Sensitive protein alignments at tree-of-life scale using
352 DIAMOND', *Nature Methods*, 18(4), pp. 366–368. Available at: [https://doi.org/10.1038/s41592-](https://doi.org/10.1038/s41592-021-01101-x)
353 [021-01101-x](https://doi.org/10.1038/s41592-021-01101-x).

354 Camacho, C. *et al.* (2009) 'BLAST+: architecture and applications', *BMC Bioinformatics*, 10(1), p. 421.
355 Available at: <https://doi.org/10.1186/1471-2105-10-421>.

356 Centers for Disease Control and Prevention (U.S.) (2019) *Antibiotic resistance threats in the United*
357 *States, 2019*. Centers for Disease Control and Prevention (U.S.). Available at:
358 <https://doi.org/10.15620/cdc:82532>.

359 Cho, J.-C. and Giovannoni, S.J. (2004) 'Cultivation and Growth Characteristics of a Diverse Group of
360 Oligotrophic Marine *Gammaproteobacteria*', *Applied and Environmental Microbiology*, 70(1), pp.
361 432–440. Available at: <https://doi.org/10.1128/AEM.70.1.432-440.2004>.

362 Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2), p. giab008.
363 Available at: <https://doi.org/10.1093/gigascience/giab008>.

364 Dereeper, A. *et al.* (2008) 'Phylogeny.fr: robust phylogenetic analysis for the non-specialist', *Nucleic
365 Acids Research*, 36(Web Server), pp. W465–W469. Available at:
366 <https://doi.org/10.1093/nar/gkn180>.

367 Dodds, D.R. (2017) 'Antibiotic resistance: A current epilogue', *Biochemical Pharmacology*, 134, pp.
368 139–146. Available at: <https://doi.org/10.1016/j.bcp.2016.12.005>.

369 Doi, Y., Wachino, J. and Arakawa, Y. (2016) 'Aminoglycoside Resistance', *Infectious Disease Clinics
370 of North America*, 30(2), pp. 523–537. Available at: <https://doi.org/10.1016/j.idc.2016.02.011>.

371 Ewels, P. *et al.* (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single
372 report', *Bioinformatics*, 32(19), pp. 3047–3048. Available at:
373 <https://doi.org/10.1093/bioinformatics/btw354>.

374 Farrell, D.J. *et al.* (2002) 'Molecular characterization of macrolide resistance mechanisms among
375 *Streptococcus pneumoniae* and *Streptococcus pyogenes* isolated from the PROTEKT 1999-2000
376 study', *Journal of Antimicrobial Chemotherapy*, 50(suppl_2), pp. 39–47. Available at:
377 <https://doi.org/10.1093/jac/dkf806>.

378 Gastineau, R. *et al.* (2012) 'Biological Activities of Purified Marennine, the Blue Pigment Responsible
379 for the Greening of Oysters', *Journal of Agricultural and Food Chemistry*, 60(14), pp. 3599–3605.
380 Available at: <https://doi.org/10.1021/jf205004x>.

381 Gastineau, R. *et al.* (2014) 'Haslea ostrearia-like Diatoms', in *Advances in Botanical Research*. Elsevier,
382 pp. 441–465. Available at: <https://doi.org/10.1016/B978-0-12-408062-1.00015-9>.

383 Harvey, K.L. *et al.* (2019) ‘The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial
384 Pathogenesis’, *Frontiers in Microbiology*, 10, p. 2351. Available at:
385 <https://doi.org/10.3389/fmicb.2019.02351>.

386 Heuer, H. *et al.* (1997) ‘Analysis of actinomycete communities by specific amplification of genes
387 encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients’, *Applied and*
388 *Environmental Microbiology*, 63(8), pp. 3233–3241. Available at:
389 <https://doi.org/10.1128/aem.63.8.3233-3241.1997>.

390 Hyatt, D. *et al.* (2010) ‘Prodigal: prokaryotic gene recognition and translation initiation site
391 identification’, *BMC Bioinformatics*, 11(1), p. 119. Available at: [https://doi.org/10.1186/1471-2105-](https://doi.org/10.1186/1471-2105-11-119)
392 [11-119](https://doi.org/10.1186/1471-2105-11-119).

393 Kolmogorov, M. *et al.* (2020) ‘metaFlye: scalable long-read metagenome assembly using repeat graphs’,
394 *Nature Methods*, 17(11), pp. 1103–1110. Available at: <https://doi.org/10.1038/s41592-020-00971-x>.

395 La Barre, S. and S. Bates, S. (2018) *Blue Biotechnology: Production and Use of Marine Molecules*.
396 Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. Available at:
397 <https://doi.org/10.1002/9783527801718>.

398 L epinay, A. *et al.* (2018) ‘First insight on interactions between bacteria and the marine diatom *Haslea*
399 *ostrearia*: Algal growth and metabolomic fingerprinting’, *Algal Research*, 31, pp. 395–405. Available
400 at: <https://doi.org/10.1016/j.algal.2018.02.023>.

401 Manni, M. *et al.* (2021) ‘BUSCO Update: Novel and Streamlined Workflows along with Broader and
402 Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes’,
403 *Molecular Biology and Evolution*. Edited by J. Kelley, 38(10), pp. 4647–4654. Available at:
404 <https://doi.org/10.1093/molbev/msab199>.

405 McGee, L. *et al.* (2015) ‘Antibiotic Resistance of Pneumococci’, in *Streptococcus Pneumoniae*.
406 Elsevier, pp. 21–40. Available at: <https://doi.org/10.1016/B978-0-12-410530-0.00002-8>.

407 Meziti, A. *et al.* (2015) ‘Spatially uniform but temporally variable bacterioplankton in a semi-enclosed
408 coastal area’, *Systematic and Applied Microbiology*, 38(5), pp. 358–367. Available at:
409 <https://doi.org/10.1016/j.syapm.2015.04.003>.

410 Munita, J.M. and Arias, C.A. (2016) ‘Mechanisms of Antibiotic Resistance’, *Microbiology Spectrum*.
411 Edited by I.T. Kudva and Q. Zhang, 4(2), p. 4.2.15. Available at:
412 <https://doi.org/10.1128/microbiolspec.VMBF-0016-2015>.

413 Murray, C.J.L. *et al.* (2022) ‘Global burden of bacterial antimicrobial resistance in 2019: a systematic
414 analysis’, *The Lancet*, 399(10325), pp. 629–655. Available at: [https://doi.org/10.1016/S0140-
415 6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).

416 Olsthoorn-Tieleman, L.N. *et al.* (2007) ‘Elongation Factor Tu3 (EF-Tu3) from the Kirromycin Producer
417 *Streptomyces ramocissimus* Is Resistant to Three Classes of EF-Tu-Specific Inhibitors’, *Journal of
418 Bacteriology*, 189(9), pp. 3581–3590. Available at: <https://doi.org/10.1128/JB.01810-06>.

419 Peticca, A., Fodil, M., Gateau, H., *et al.* (2023a) ‘Complete mitochondrial genome and draft
420 chloroplastic genome of *Haslea ostrearia* (Naviculaceae, Bacillariophyceae) [Manuscript submitted
421 for publication]’. Available at : <https://doi.org/10.1101/2022.11.09.515767>.

422 Peticca, A., Fodil, M., Chenaïs, B., *et al.* (2023b) ‘The need to stabilize the bacterial community present
423 in the cultures of the marine diatom *Haslea ostrearia* [Manuscript submitted for publication]’.
424 Available at : <http://dx.doi.org/10.21203/rs.3.rs-2751923/v1> . Pouvreau, J.-B. *et al.* (2008)
425 ‘Antioxidant and Free Radical Scavenging Properties of Marennine, a Blue-Green Polyphenolic
426 Pigment from the Diatom *Haslea ostrearia* (Gaillon/Bory) Simonsen Responsible for the Natural
427 Greening of Cultured Oysters’, *Journal of Agricultural and Food Chemistry*, 56(15), pp. 6278–6286.
428 Available at: <https://doi.org/10.1021/jf073187n>.

429 PubChem (no date a) *Amphotericin B Compound Summary*. Available at:
430 <https://pubchem.ncbi.nlm.nih.gov/compound/5386092> (Accessed: 24 May 2023).

431 PubChem (no date b) *Streptomycin Compound Summary*. Available at:
432 <https://pubchem.ncbi.nlm.nih.gov/compound/19649> (Accessed: 24 May 2023).

433 Quast, C. *et al.* (2012) ‘The SILVA ribosomal RNA gene database project: improved data processing
434 and web-based tools’, *Nucleic Acids Research*, 41(D1), pp. D590–D596. Available at:
435 <https://doi.org/10.1093/nar/gks1219>.

436 Quinlan, A.R. and Hall, I.M. (2010) ‘BEDTools: a flexible suite of utilities for comparing genomic
437 features’, *Bioinformatics*, 26(6), pp. 841–842. Available at:
438 <https://doi.org/10.1093/bioinformatics/btq033>.

439 Seemann, T. (2014) ‘Prokka: rapid prokaryotic genome annotation’, *Bioinformatics*, 30(14), pp. 2068–
440 2069. Available at: <https://doi.org/10.1093/bioinformatics/btu153>.

441 Selvarajan, R. *et al.* (2022) ‘Evolution and Emergence of Antibiotic Resistance in Given Ecosystems:
442 Possible Strategies for Addressing the Challenge of Antibiotic Resistance’, *Antibiotics*, 12(1), p. 28.
443 Available at: <https://doi.org/10.3390/antibiotics12010028>.

444 Simonsen, R. (1974) *The Diatom Plankton of the Indian Ocean Expedition of RV ‘Meteor’ 1964-1965*.
445 Gebrüder Borntraeger (Biology, 19).

446 *The Comprehensive Antibiotic Resistance Database* (no date). Available at:
447 <https://card.mcmaster.ca/analyze/rgi> (Accessed: 24 May 2023).

448 Turpin, V. *et al.* (2001) ‘Oyster greening by outdoor mass culture of the diatom *Haslea ostrearia*
449 Simonsen in enriched seawater: Mass production of the diatom *Haslea ostrearia*’, *Aquaculture*
450 *Research*, 32(10), pp. 801–809. Available at: <https://doi.org/10.1046/j.1365-2109.2001.00615.x>.

451 Vaser, R. *et al.* (2017) ‘Fast and accurate de novo genome assembly from long uncorrected reads’,
452 *Genome Research*, 27(5), pp. 737–746. Available at: <https://doi.org/10.1101/gr.214270.116>.

453 Vasimuddin, Md. *et al.* (2019) ‘Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore
454 Systems’, in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*.

455 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Rio de Janeiro,
456 Brazil: IEEE, pp. 314–324. Available at: <https://doi.org/10.1109/IPDPS.2019.00041>.

457 Wireman, J. *et al.* (1997) ‘Association of mercury resistance with antibiotic resistance in the gram-
458 negative fecal bacteria of primates’, *Applied and Environmental Microbiology*, 63(11), pp. 4494–
459 4503. Available at: <https://doi.org/10.1128/aem.63.11.4494-4503.1997>.

460 Wyres, K.L. *et al.* (2013) ‘Evidence of antimicrobial resistance-conferring genetic elements among
461 pneumococci isolated prior to 1974’, *BMC Genomics*, 14(1), p. 500. Available at:
462 <https://doi.org/10.1186/1471-2164-14-500>.

463 Zhang, X.-X., Zhang, T. and Fang, H.H.P. (2009) ‘Antibiotic resistance genes in water environment’,
464 *Applied Microbiology and Biotechnology*, 82(3), pp. 397–414. Available at:
465 <https://doi.org/10.1007/s00253-008-1829-z>.

466

Cet article met en avant la présence, dans chaque bactérie, d'au moins une résistance aux antibiotiques actuellement utilisés dans le protocole de culture. Ces antibiotiques à large spectre ont précédemment été soupçonnés de n'avoir que peu d'effet sur cette communauté bactérienne, et les observations faites ici semblent le confirmer. En effet, beaucoup de génomes bactériens différents ont été retrouvés dans le métagénome, dont plus d'une trentaine quasiment entiers et avec une grande profondeur de séquençage. Ces deux observations indiquent que leur présence dans les cultures était importante, bien plus que celui de *H. ostrearia* qui n'a pu être assemblé (Chapitre 1). Mais cette fois-ci, l'étude a permis de proposer un premier élément de réponse à la présence de ces nombreuses bactéries malgré le traitement antibiotique : la présence de gènes de résistances ARGs et AMRs. Il a été démontré, une nouvelle fois, la nécessité d'utiliser des antibiotiques spécifiques aux bactéries présentes dans les cultures pour réduire la concentration et/ou la diversité de ces bactéries. Une liste des antibiotiques à éviter ou à tester en priorité a ainsi été proposée en tenant compte des résistances retrouvées pour chaque bactérie. D'ailleurs, des analyses devraient être réalisées par la doctorante Amel Benali Khitem dans notre laboratoire pour vérifier l'effet de nombreux antibiotiques sur les cultures de *H. ostrearia*, mais également sur les bactéries seules (sous forme d'antibiogrammes).

De plus, comme présenté dans la figure 2 de l'article, les ARGs détectés sont très diversifiés. De nombreuses résistances ont été retrouvées dans chacune des bactéries les mieux assemblées, et donc les plus présentes dans les cultures, dont certaines spécifiques à une famille bactérienne. D'après ces résultats, il serait donc nécessaire de réaliser un cocktail d'antibiotiques pour pouvoir éliminer un grand nombre de bactéries des cultures de *H. ostrearia*. De plus, l'utilisation de plusieurs antibiotiques spécifiques permettrait d'analyser de façon ciblée, en supprimant spécifiquement un antibiotique du cocktail, quelles sont les bactéries nécessaires à la prolifération des microalgues. En effet, ces dernières ne présentent pas les mêmes résistances. Avec des antibiotiques capables de n'éliminer qu'un nombre réduit d'organismes, il serait plus facile de croiser les résultats obtenus et d'identifier les bactéries potentiellement nécessaires à la survie de *H. ostrearia*. D'autant que ces gènes de résistance

ne semblent pas être issus d'une mutation récente mais d'un trait transmis par un ancêtre lointain, cette conclusion a été notamment démontrée pour les gènes modifiés *bcr* et *EF-Tu*. Les résistances observées pour chaque bactérie ne seraient donc pas spécifiques à nos cultures, mais seraient également observables en milieu naturel. Toutes les cultures réalisées à partir d'un des isolats faits à Bouin en 2018 pourraient donc profiter d'un nouveau protocole de culture amélioré, à condition que les mêmes taxons bactériens y soient retrouvés (Poulin et al 2019).

En dernier lieu, l'existence de résistances à des antibiotiques classés comme "dernier recours" a été notifiée. Comme leurs noms l'indiquent, des contrôles s'appliquent pour obtenir et utiliser ces antibiotiques : soit parce qu'ils sont particulièrement générateurs de résistances bactériennes, soit parce qu'ils représentent un intérêt particulier dans le cas d'un traitement médical (dernier antibiotique efficace contre une bactérie donnée par exemple). Dans le contexte de nos bactéries, ces résistances ne semblent pas représenter un danger immédiat. Cependant, ce caractère inquiétant se renforce si nous changeons de perspective. En effet, bien que les bactéries actuellement identifiées ne soient pas pathogènes, il n'en reste pas moins qu'elles ont été isolées en même temps que *H. ostrearia*. Or, cette dernière a été récupérée dans des bassins ostréicoles et est souvent filtrée par les huîtres comme le prouve la couleur verdâtre de leurs branchies (Gastineau et al 2012). Il est donc fort probable que les bactéries associées à nos cultures en fassent de même. Ces bivalves pourraient ensuite être mangés, crus, par un être humain, ce qui pourrait permettre aux bactéries marines d'entrer en contact avec le microbiote humain et/ou des bactéries et champignons pathogènes (Salmonelles, *Candida*, etc), et potentiellement leur transmettre des résistances aux antibiotiques de dernier recours. On estime que plus de 2,8 millions d'infections par an sont causées par des bactéries ou des champignons résistants aux antibiotiques rien qu'aux États-Unis, entraînant 35 900 décès (Centers for Disease Control and Prevention (U.S.) 2019). Néanmoins, la question des AMRs bactériens ne se limite pas à cette seule région, puisqu'elle s'est révélée être un problème de santé majeur à l'échelle mondiale (Murray et al 2022). A cela s'ajoutent des problèmes de santé graves entraînés par la colonisation d'un environnement Y

par des bactéries impossible à éliminer au vu de leurs multiples résistances, alors qu'elles sont parfaitement inoffensives dans un environnement X. Ce cas a déjà été rapporté pour *Pseudomonas aeruginosa*, inoffensive dans l'intestin, mais qui s'est avérée mortelle en grande concentration dans les poumons, car résistante à tous les antibiotiques disponibles (Wheatley et al 2022).

Pour toutes ces raisons, il est particulièrement intéressant d'étudier les résistances retrouvées chez ce microbiote que cela soit d'un point de vue d'amélioration des cultures ou de compréhension de l'environnement marin.

3. Résultats complémentaires

Dans l'objectif de vérifier la facilité de transmission de ces résistances à d'autres bactéries, la présence des AMRs ou ARGs a été vérifiée dans les plasmides bactériens contenus dans le métagénome. En effet, la transmission horizontale de gènes entre bactéries s'avère plus difficile lorsque ces derniers se situent dans le génome nucléaire (Chapitre 3-2), ce qui n'est pas le cas pour des gènes localisés dans les plasmides. Le mécanisme le plus répandu pour ce type de transfert d'ADN est la conjugaison bactérienne (Norman et al 2009), mécanisme qui repose sur les plasmides. Il n'est donc que plus facile d'acquérir des résistances si ces dernières sont issues de gènes plasmidiques.

Les génomes plasmidiques ont donc été recherchés dans le métagénome à l'aide du logiciel Platon 1.6 (Schwengers et al 2020). Au total, 21 contigs ont été catégorisés comme plasmides (Tableau 3-1), et sont toutes des séquences supposées partielles, car non circularisées ainsi que nouvellement identifiées comme des séquences bactériennes, à part le contig 646 qui a déjà été identifié auparavant comme génome bactérien grâce à la recherche d'ARNr 16S (Chapitre 1-2). Toutefois, le score BUSCO associé à ce dernier ne dépassait pas les 50%, et il n'a pas été analysé plus avant. Cependant, afin de s'assurer que Platon n'aurait pas commis d'erreur en l'annotant comme plasmide, sa séquence a été soumise à un Blastn 2.14.0 contre la base de données nt du NCBI. Malheureusement aucun résultat n'a pu être retrouvé et les résultats obtenus pour le contig 2646 seront donc à prendre avec beaucoup de précaution par la suite. La même méthodologie a été employée pour annoter taxonomiquement les autres contigs, les résultats de la recherche par Blastn sur la base de données nt ont été filtrés pour ne garder que le taxon associé à l'alignement le plus long avec un pourcentage d'identité (%ID) supérieur à 90%. Toutes les séquences ayant été alignées sur un plasmide bactérien ont ainsi pu être associées à un rang taxonomique. Et, comme pour l'ensemble des analyses précédentes, les résultats obtenus indiquent que ces plasmides proviennent de bactéries marines, il est donc fortement probable que ces dernières proviennent également de l'environnement naturel de *H. ostrearia* et non d'une contamination

des cultures. Cependant, 5 contigs ne partagent aucune correspondance avec nt, et n'ont donc pas été identifiés. Ces derniers appartiennent aux 6 séquences les plus longues (>200kb), avec celle du contig 2646, sans qu'aucune explication ne puisse être donnée concernant cette observation. Le contig 2729 a également été associé au genre des *Sagittula* mais l'alignement n'atteint que 81.711%ID, il a donc été marqué comme incertain. Une tentative d'améliorer les résultats obtenus a été testée grâce au logiciel COPLA du 30 juin 2021 (Abby et al 2016), spécialisé dans la classification taxonomique de plasmide, mais en vain. Les gènes de résistance ont quant à eux été recherchés avec les techniques listées dans l'article AMRs et ARGs (Chapitre 3-2) : Prokka et CARD vs RGI.

Seules quelques résistances ont pu être retrouvées grâce au logiciel Prokka. Les génomes plasmidiques associées à ces dernières ont été dans les plus grands identifiées (>200kb), à l'exception du contig 3618 qui n'a atteint qu'une taille de 191,255b. La moitié des résistances provenaient de modification des gènes *MdtA* et *MexB*, eux-mêmes associées à un mécanisme de multirésistance par pompe à efflux. Pour toutes les raisons évoquées dans l'article précédent (Chapitre 3-2), ils ont donc été ignorés pour la suite. Les gènes les plus intéressants retrouvés dans nos données ont été *CzcA* et *CzcB*, *Acr3* et un opéron de protéine de régulation. Ces derniers sont associés, respectivement, à une résistance contre le cobalt-zinc-cadmium, à l'arsenic et au mercure. Pour *Acr3*, ce gène a été retrouvé dans un plasmide de *Sulfitobacter*, or un génome entier, ou quasiment, a déjà été retrouvé pour ce genre bactérien dans les données (contig 3360 avec 78.9%BUSCO et circulaire, Chapitre 3-2). Ce dernier possédait 14 ARGs différents, dont 8 non associés à une multirésistance, mais aucun à l'arsenic comme ici. Il s'agirait donc d'une nouvelle résistance à ajouter à ce type bactérien. Cependant il est impossible de savoir si ce plasmide et le génome récupérés sont liés, en d'autres termes présents dans la même cellule. Ces derniers pourraient être issus d'espèces différentes ou simplement d'individus différents, rien ne nous permet ici de le vérifier. D'autres résistances ont été retrouvées dans le contig 2646, notamment celle *YaaA* qui confère une résistance au stress oxydant. Mais devant l'incapacité de confirmer qu'il s'agit bien d'un plasmide, les résultats ne sont donnés qu'à titre d'information. Le génome entier d'un

Oceanicaulis a été retrouvé dans les données (contig 3479 avec 98.1%BUSCO et circulaire, Chapitre 3-2), mais avec d'autres gènes que *YaaA* liés à l'oxydation (*oxyR*, *ahpD*) et plus d'une dizaine de ARGs différents.

Devant ces maigres résultats, l'inquiétude sur la transmission de gènes de résistance à des antibiotiques, de dernier recours ou fortement utilisés, ne semble pas fondée. Bien des résistances aient été retrouvées, aucune de ces dernières ne sont associées à des antibiotiques et la très grande majorité des génomes plasmidiques (17/21) n'en possèdent pas.

Contig	Longueur (bases)	Circulaire ?	%GC	Taxon identifié par NCBI	Résistances retrouvées par Prokka
contig_2133	444539	non	57,78%	N/A	Multidrug MdtA/MexB
contig_3596	409587	oui	55,12%	N/A	N/A
contig_2959	271847	non	57,17%	N/A	Multidrug MexB/MdtA
contig_4666	244758	non	54,75%	N/A	Cobalt-zinc-cadmium resistance protein CzcA/CzcB, Mercuric resistance operon regulatory protein x2
contig_2646	225181	non	50,89%	<i>Oceanicaulis</i>	Peroxide stress resistance protein YaaA, Multidrug MdtE/MexB
contig_4685	208823	non	57,58%	N/A	N/A
contig_3618	191255	non	56,21%	<i>Sulfitobacter</i>	Arsenical-resistance protein Acr3
contig_3594	188524	non	56,52%	<i>Rhodobacterales</i>	N/A
contig_1916	179960	non	59,55%	<i>Roseobacter</i>	N/A
contig_7755	107772	non	66,14%	<i>Antarctobacter heliothermus</i>	N/A
contig_1914	89539	non	57,02%	<i>Roseobacter</i>	N/A
contig_3126	86327	non	61,55%	<i>Roseobacteraceae</i>	N/A
contig_2284	73990	non	56,81%	<i>Salipiger</i>	N/A
contig_3615	67358	non	59,82%	<i>Roseobacteraceae</i>	N/A
contig_3832	58586	non	59,17%	<i>Rhodobacterales</i>	N/A
contig_3616	54025	non	57,78%	<i>Roseobacteraceae</i>	N/A
contig_1991	48063	non	58,08%	<i>Sulfitobacter</i>	N/A
contig_2963	46517	non	49,29%	<i>Paracoccaceae</i>	N/A
contig_3945	45735	non	54,90%	<i>Roseobacteraceae</i>	N/A
contig_2729	32745	non	58,17%	<i>Sagittula</i>	N/A
contig_2962	31852	non	58,41%	<i>Paracoccaceae</i>	N/A

Tableau 3-1 : Informations sur les plasmides identifiées dans le métagénome obtenu pour les cultures de *Haslea ostrearia* et les gènes de résistance aux antibiotiques ou aux antimicrobiens retrouvés à l'intérieur de leur séquence

Les plasmides ont été identifiés grâce au logiciel Platon 1.6 (Schwengers et al 2020), et les gènes de résistance par Prokka 1.14.6 (Seemann 2014). Aucun résultat n'a été obtenu par l'analyse de CARD (Alcock et al 2023) 3.2.5 et RGI 6.0.1 (Alcock et al 2020).

Une identification taxonomique est proposée grâce à un Blastn (Altschul et al 1990) contre la base de données nt du NCBI. Si une séquence homologue est retrouvée avec un pourcentage de similarité de séquences (%id) supérieur à 90%, alors le contig est annoté comme faisant partie du même taxon.

4. Conclusion du chapitre

De nombreuses résistances ont été retrouvées dans les génomes bactériens nucléaires, et quelques-unes dans ceux plasmidiques. Parmi elles se trouvent des insensibilités à certains antibiotiques fréquemment retrouvées chez les bactéries, notamment suite à leur mauvais usage jusque dans les années 2000 (Barber et Rozwadowska-Dowzenko 1948, Jevons 1961, Teuber 2001, Aminov 2009), ce qui expliquerait l'absence d'efficacité du traitement antibiotique actuellement utilisé pour les cultures de *H. ostrearia*.

Dans leur totalité, les AMRs retrouvés sont très variés, et proviennent d'espèces bactériennes diverses. Il semblerait également qu'ils n'aient pas été nouvellement acquis, et ne soient donc pas issus d'une mutation engendrée par le protocole de culture. Cette dernière hypothèse est d'ailleurs soutenue par la faible présence d'AMRs dans les plasmides. Il y a donc peu de chance que des transmissions horizontales aient eu lieu récemment.

Également, il a été retrouvé des gènes conférants des résistances à certains antibiotiques de dernier recours. Bien qu'actuellement peu dangereuse, cette présence pourrait être inquiétante dans le futur, notamment si ces ARGs venaient à être transmis à d'autres bactéries pathogènes.

Ces résultats sont intéressants du point de vue du laboratoire. L'utilisation de ces nouvelles connaissances sur la grande diversité des AMRs, et leur présence dans seulement certains génomes, permettrait de dresser une première liste des bactéries potentiellement nécessaires à la survie de *H. ostrearia*. Une fois les bactéries potentiellement nécessaires identifiées, il serait également possible de vérifier si celles-ci sont retrouvées pour n'importe quel isolat de la microalgue (en conditions naturelles). Il serait aussi intéressant de vérifier si ces dernières partagent les mêmes AMRs au sein d'une même population (signe d'une évolution commune).

D'un point de vue écologique, ces résistances sont aussi intrigantes. Comment ces résistances ont pu être retrouvées dans des bactéries marines évoluant dans un bassin

ostréicole ? D'où proviennent leurs insensibilités aux antibiotiques de dernier recours ? Il serait intéressant de vérifier si elles sont également retrouvées chez d'autres bactéries du même site, et pas forcément chez celles associées à *H. ostrearia*.

Conclusion générale

“Les choses ne se font pas toujours comme ça, nettement pour des raisons précises. Elles se tassent, se glissent, elles arrivent d'elles-mêmes et souvent sans qu'on le veuille à une conclusion qui se fait peu à peu accepter.” - extrait de Hervé Bazin, *Qui j'ose aimer* (1956)

1. Remise en contexte des résultats

Haslea ostrearia est une espèce de microalgue peu connue et dont l'étude peut présenter de nombreux obstacles: co-culture bactérienne obligatoire, protocole complexe... Si initialement le projet de thèse devait se focaliser sur l'étude de la structure du génome de *H. ostrearia*, ces différentes problématiques ont amené à prendre une nouvelle orientation scientifique. Ainsi, l'hypothèse que de nombreuses bactéries vivent avec la microalgue et pourraient être importantes pour sa survie a servi de base à une nouvelle réflexion: toutes les bactéries présentes en culture sont-elles nécessaires? Si non, comment identifier celles essentielles ?

Au moment de l'écriture de cette thèse, il n'existe que très peu de connaissances sur la génétique de *H. ostrearia*, à l'instar des autres diatomées. En effet, en dehors des travaux exposés ici, seuls les séquences de ses plasmides ont été dévoilés récemment. Ces derniers ont fait l'objet d'une publication deux ans après les assemblages présentés lors du Chapitre 1 (Dąbek et al 2022). Il semblerait que, tout comme les résultats obtenus précédemment, cet article n'ait pas pu reconstruire le génome nucléaire et se soit concentré sur celui des organelles. Même avec l'utilisation d'une autre souche de *H. ostrearia* et des temps de culture différents, il semble être toujours complexe d'obtenir l'ADN complet de cette microalgue. Le même problème s'est fait ressentir lors de leurs analyses: des bactéries contaminent les cultures et par extension les séquençages ADN. De plus, de précédentes analyses publiées

décrivent un effet positif potentiel de ces dernières sur la microalgue (Lépinay et al 2016, Lépinay et al 2018), mais sans jamais proposer de mécanisme. Les interactions entre *H. ostrearia* et son environnement naturel ne sont également pas documentées, contrairement à d'autres diatomées dont on suppose qu'elles entretiennent une relation mutualiste avec certaines bactéries (Jauffrais et al 2017, Koedooder et al 2019, Shibl et al 2020, Sterling et al 2023).

Devant ces faits, il est donc apparu important de considérer la population bactérienne comme une composante essentielle des cultures de *H. ostrearia*, et de l'intégrer dans les protocoles de laboratoire. Il est courant que les cultures de *H. ostrearia* ne soient que très peu, voire pas du tout traitées par des antibiotiques (Nghiem-Xuan 2019, Nghiem Xuan et al 2020, Adjout et al 2022), les protocoles expérimentaux récents n'utilisant que des antibiotiques à large spectre. Or, lors de nos travaux du Chapitre 2, il a été mis en évidence que leurs impacts sur les bactéries présentes dans les cultures ne sont pas suffisants, bien qu'ils semblent aider à stabiliser ces dernières (concentrations bactériennes stables entre les cultures). Contrôler les bactéries présentes dans les cultures permettrait ainsi de réduire leur diversité, sans impacter la croissance de *H. ostrearia*, mais également d'apporter une reproductibilité des cultures (Jones et al 1973). En ne se concentrant pas seulement sur l'identification de ces dernières mais également sur les gènes de résistances leur étant associés, les résultats présentés dans le Chapitre 3 permettent d'établir les antibiotiques n'ayant peu ou pas d'effet sur une liste de bactéries connues et aider dans l'établissement de ce nouveau protocole.

L'ensemble des analyses repose sur des données métagénomiques, c'est-à-dire que l'ensemble des ADN présents dans les cultures ont été séquencés. Avec la baisse des coûts de séquençage, cette méthode est aujourd'hui très intéressante et favorise l'étude de microbiomes (Bharti et Grimm 2021). Auparavant les identifications bactériennes ne reposaient que sur l'étude de morceaux du gène ARN 16S, qui présentent des inconvénients importants tels que l'incapacité de calculer l'abondance d'une espèce ou les résultats variants selon la région sélectionnée. Les travaux présentés lors du Chapitre 2 en sont d'ailleurs la preuve. Reconstituer des génomes entiers à partir de séquençages métagénomiques sans

passer par une étape d'amplification de l'ADN permet de pallier ces problèmes (Bharti et Grimm 2021). De plus, les logiciels d'analyse de métagénomiques s'améliorent grandement depuis les dernières années: de nombreux logiciels sont développés pour identifier les espèces (West et al 2018, Wood et al 2019), ou reconstituer des génomes de très bonne qualité (Kolmogorov et al 2020, Latorre-Pérez et al 2020).

2. La place de la métagénomique dans l'amélioration et la compréhension des manipulations biologiques

Comme précédemment évoqué, l'utilisation des données métagénomiques s'est développée grandement depuis ces dernières années. Aujourd'hui, cette méthode de séquençage est même qualifiée comme le moyen le "plus efficace pour détecter des agents pathogènes nouveaux, rares et inattendus et pour traiter les co-infections en comparaison avec les méthodes de diagnostic conventionnelles" (*"Compared with conventional diagnostic methods, metagenomic sequencing is better in detecting novel, rare, and unexpected pathogens and being applied in co-infections."* Chen et al 2022).

En effet, contrairement aux séquençages d'une espèce unique, la métagénomique permet d'obtenir l'ensemble des informations génétiques d'espèces non cultivables seules et/ou d'espèces inconnues. Ces espèces ne sont pas obligatoirement bactériennes comme avec le séquençage 16S, mais peuvent appartenir à n'importe quel type d'organismes vivants (bactéries, virus, champignons, eucaryotes, etc.). Également, les données métagénomiques permettent d'obtenir des informations fonctionnelles (fonctions des gènes) sur un ensemble de génomes, informations inaccessibles par d'autres moyens de séquençage. Des études sur le microbiome humain suggèrent que ces informations permettent d'identifier efficacement les différences entre des microbiomes sains et malades (Martín et al 2014, Syama et al 2023), mais aussi d'identifier avec précision des gènes conférant des résistances aux antibiotiques (Yorki et al 2023). Et ce, malgré l'une des plus grands limitations dont nous avons fait état

précédemment, l'identification des individus lorsque ceux-ci sont trop nombreux ou inconnus, que de nombreux logiciels tentent d'améliorer (West et al 2018, Wood et al 2019, Jiang et al 2023, Lataretu et al 2023).

Dans notre cas, l'utilisation d'un métagénome a permis de répondre à plusieurs questions qui se sont posées au fur et à mesure de l'avancement des travaux :

- D'où proviennent les bactéries vivant dans les cultures de *H. ostrearia* ?
- Combien de bactéries différentes sont présentes dans les données ?
- Pourquoi cette contamination bactérienne est toujours observable même après un traitement antibiotique ? Ce dernier a-t-il un effet sur la population bactérienne ?

Ainsi, grâce aux génomes qui ont été assemblés à partir des données métagénomiques, il a pu être compris que ces bactéries proviennent toutes de milieux marins et possèdent des résistances aux antibiotiques/antimicrobiens. Leur abondance a également permis de comprendre l'utilité de trouver un traitement adapté à ces dernières pour contrôler la population bactérienne.

Les analyses métagénomiques réalisés au cours de cette thèse permettent d'orienter les manipulations en laboratoire actuelles sur *H. ostrearia* en justifiant la présence des bactéries.

3. Perspectives proches et lointaines

La perspective proche envisagée est celle pour laquelle les recherches sur la diversité bactérienne et leurs caractéristiques ont été réalisées : l'amélioration des protocoles de culture de *H. ostrearia*. Suite à l'identification des antibiotiques avec un impact sur certains groupes de bactéries, il est aujourd'hui possible de tester ces derniers en laboratoire pour connaître les bactéries potentiellement nécessaires à la croissance de la microalgue. En notant l'effet sur cette dernière pour chaque antibiotique et en croisant les résultats obtenus, il pourrait être

envisagé d'obtenir une liste réduite de bactéries à conserver ou à éliminer. Il serait également intéressant de coupler ces analyses à un séquençage 16S pour vérifier les bactéries encore présentes dans les cultures et valider le bon déroulement des tests : Est-ce qu'il s'agit des bactéries attendues ? Lesquelles ont disparu entre ces tests et celles identifiées ici ? Bien que moins précis face aux analyses métagénomiques réalisées ici, le séquençage 16S reste plus facile d'utilisation pour des biologistes classiques et à un coût moins élevé.

Des tests avec des antibiotiques pour lesquels aucune résistance n'a été retrouvée dans les génomes bactériens pourront également être réalisés, nous informant ainsi sur la nécessité ou non des bactéries pour la survie de *H. ostrearia*. Ces derniers pourraient également être importants pour comprendre les conditions à réunir pour la production de marennine: ce pigment n'est-il produit que lorsque des bactéries sont présentes dans l'environnement ? Si oui, il sera alors nécessaire de réaliser d'autres expérimentations pour comprendre si la production de marennine est déclenchée par leur présence ou issue d'une relation symbiotique entre les bactéries et la microalgue.

Aussi, une fois que l'identité des bactéries "essentiels" à la croissance de *H. ostrearia* connue, il pourrait être envisagé de rechercher ces dernières dans les autres milieux où la microalgue est retrouvée. Grâce aux génomes de ces bactéries "essentiels" qui ont été reconstitués précédemment, il serait possible de rechercher leur présence dans des séquençages des isolats de *H. ostrearia*, mais également d'explorer les bases de données constituées d'ADN récupérés en milieu marin. Il faudra cependant vérifier au préalable la présence de la microalgue dans les échantillons grâce aux séquences de la mitochondrie, du chloroplaste ou de celles identifiées comme appartenant au génome nucléaire. Cette analyse permettra d'accentuer le caractère essentiel de ces bactéries à la survie de *H. ostrearia* si ces dernières sont systématiquement retrouvées avec elle, et la potentielle mise en évidence de bactéries évoluant selon les conditions environnementales. Par exemple, est-ce que l'on dénote des bactéries aidant les organismes environnants à résister à des températures glaciales dans des eaux plus froides ? Il serait également important d'effectuer plusieurs prélèvements sur une même population mais à des temps différents qui correspondent à

différents stades dans le cycle de vie de *H. ostrearia*. Leurs analyses pourraient aider à observer et comprendre les évolutions potentielles dans la composition bactérienne des colonies de *H. ostrearia* au fil du temps.

Ensuite, lorsque le génome de *H. ostrearia* sera d'assez bonne qualité pour être étudié et annoté, les génomes bactériens présentés ici pourront être analysés pour comprendre leurs apports pour la microalgue. Par exemple, des gènes de production/assimilation de certains nutriments nécessaires pour *H.ostrearia* pourraient être absents de son génome mais retrouvés dans celui de bactéries vivant avec elle. Ce phénomène pourrait indiquer une évolution commune entre les deux espèces, et donc une symbiose obligatoire entre ces dernières. Des symbioses facultatives pourraient également être observées avec des gènes bactériens qui peuvent stimuler l'expression de gènes de la microalgue et augmenter sa croissance. Pour identifier les gènes nécessaires à la survie de *H. ostrearia*, une liste de ces derniers devra être dressée en s'appuyant sur les connaissances actuelles du genre mais aussi des diatomées. Pour les gènes bactériens stimulant ou apportant des nutriments supplémentaires à la microalgue, il faudra croiser les informations métabolomiques avec les besoins de *H. ostrearia*.

Si l'on souhaite également se pencher plus sur le sujet de la transmission des ARGs et AMRs, une recherche de bactériophages (phages) dans les cultures de *H. ostrearia* pourrait être envisagée. Ces derniers peuvent contenir des gènes accessoires à leur cycle de vie dans leur génome, et participer ainsi aux transferts horizontaux de gènes entre bactéries de la même espèce (Moon et al 2015, Touchon et al 2017). Bien que l'infection d'une bactérie par un phage est un mécanisme précis, ce processus peut varier dans de rares cas et engendrer des conséquences directes sur cette dernière. Il est ainsi tout à fait possible que des gènes bactériens se glissent dans la capsid lorsque le matériel génétique du phage y est empaqueté, et soient transmis à une autre bactérie lors d'une prochaine infection par ce même bactériophage.

Bibliographie

- Abby, S.S. et al. (2016) 'Identification of protein secretion systems in bacterial genomes', *Scientific Reports*, 6(1), p. 23080. Disponible sur: <https://doi.org/10.1038/srep23080>.
- Adjout, R. et al. (2022) 'Effects of temperature, irradiance, and pH on the growth and biochemical composition of *Haslea ostrearia* batch-cultured in an airlift plan-photobioreactor', *Applied Microbiology and Biotechnology*, 106(13–16), pp. 5233–5247. Disponible sur: <https://doi.org/10.1007/s00253-022-12055-1>.
- Adl, S.M. et al. (2012) 'The Revised Classification of Eukaryotes', *Journal of Eukaryotic Microbiology*, 59(5), pp. 429–514. Disponible sur: <https://doi.org/10.1111/j.1550-7408.2012.00644.x>.
- Alcock, B.P. et al. (2019) 'CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database', *Nucleic Acids Research*, p. gkz935. Disponible sur: <https://doi.org/10.1093/nar/gkz935>.
- Alcock, B.P. et al. (2023) 'CARD 2023: expanded curation, support for machine learning, and resistance prediction at the Comprehensive Antibiotic Resistance Database', *Nucleic Acids Research*, 51(D1), pp. D690–D699. Disponible sur: <https://doi.org/10.1093/nar/gkac920>.
- Ale, M.T. and Meyer, A.S. (2013) 'Fucoidans from brown seaweeds: An update on structures, extraction techniques and use of enzymes as tools for structural elucidation', *Rsc Advances*, 3(22), pp. 8131–8141.
- Ali, H., Al-Darraj, M. and Ali, S. (2021) 'The Allelopathic in the Diatoms', *Rafidain Journal of Science*, 30(3), pp. 30–38. Disponible sur: <https://doi.org/10.33899/rjs.2021.168909>.
- Allen, A.E. et al. (2011) 'Evolution and metabolic significance of the urea cycle in photosynthetic diatoms', *Nature*, 473(7346), pp. 203–207. Disponible sur: <https://doi.org/10.1038/nature10074>.
- Allen, A.E., Vardi, A. and Bowler, C. (2006) 'An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms', *Current opinion in plant biology*, 9(3), pp. 264–273.
- Altschul, S.F. et al. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. Disponible sur: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Alverson, A.J. et al. (2011) 'The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*', *BMC evolutionary biology*, 11(1), pp. 1–8.

- Amin, S.A. et al. (2015) 'Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria', *Nature*, 522(7554), pp. 98–101. Disponible sur: <https://doi.org/10.1038/nature14488>.
- Aminov, R.I. (2009) 'The role of antibiotics and antibiotic resistance in nature', *Environmental Microbiology*, 11(12), pp. 2970–2988. Disponible sur: <https://doi.org/10.1111/j.1462-2920.2009.01972.x>.
- Aminov, R.I. (2010) 'A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future', *Frontiers in Microbiology*, 1. Disponible sur: <https://doi.org/10.3389/fmicb.2010.00134>.
- Aminov, R.I. and Mackie, R.I. (2007) 'Evolution and ecology of antibiotic resistance genes', *FEMS microbiology letters*, 271(2), pp. 147–161.
- Andrews, S. (2010) 'FastQC: A Quality Control Tool for High Throughput Sequence Data'. Disponible sur: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anjuli, B. (2019) Communicating Antibiotic Resistance to the Public: How effective was Public Health England's 2018 'Keep Antibiotics Working' campaign TV advertisement at increasing public understanding of antibiotic resistance and motivating a change in antibiotic seeking behaviours? Malmö universitet/Kultur och samhälle.
- Armbrust, E.V. et al. (2004) 'The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism', *Science*, 306(5693), pp. 79–86. Disponible sur: <https://doi.org/10.1126/science.1101156>.
- Armbrust, E.V. (2009) 'The life of diatoms in the world's oceans', *Nature*, 459(7244), pp. 185–192.
- Asagabaldan, M.A. et al. (2023) 'First Record of Epizoic Diatom Presence on *Poricellaria ratoniensis* (Bryozoa, Cheilostomata) from Java Sea, Indonesia', *Journal of Science and Applicative Technology*, 7(1), pp. 52–56.
- Avery, O.T. and Dubos, R. (1930) 'The specific action of a bacterial enzyme on pneumococci of type III', *Science*, 72(1858), pp. 151–152.
- Bankevich, A. et al. (2012) 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing', *Journal of computational biology*, 19(5), pp. 455–477.
- Barber, M. and Rozwadowska-Dowzenko, M. (1948) 'INFECTION BY PENICILLIN-RESISTANT STAPHYLOCOCCI', *The Lancet*, 252(6530), pp. 641–644. Disponible sur: [https://doi.org/10.1016/S0140-6736\(48\)92166-7](https://doi.org/10.1016/S0140-6736(48)92166-7).
- Beckers, B. et al. (2016) 'Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies', *Frontiers in Microbiology*, 7. Disponible sur: <https://doi.org/10.3389/fmicb.2016.00650>.
- Bedoshvili, Y.D., Popkova, T.P. and Likhoshway, Y.V. (2009) 'Chloroplast structure of diatoms of different classes', *Cell and Tissue Biology*, 3, pp. 297–310.

- Behringer, G. et al. (2018) 'Bacterial Communities of Diatoms Display Strong Conservation Across Strains and Time', *Frontiers in Microbiology*, 9, p. 659. Disponible sur: <https://doi.org/10.3389/fmicb.2018.00659>.
- Bélanger, W. et al. (2020) 'Extraction Improvement of the Bioactive Blue-Green Pigment "Marennine" from Diatom *Haslea ostrearia*'s Blue Water: A Solid-Phase Method Based on Graphitic Matrices', *Marine Drugs*, 18(12), p. 653. Disponible sur: <https://doi.org/10.3390/md18120653>.
- Ben Ali, A. et al. (2001) 'Phylogenetic relationships among algae based on complete large-subunit rRNA sequences.', *International journal of systematic and evolutionary microbiology*, 51(3), pp. 737–749.
- Bharti, R. and Grimm, D.G. (2021) 'Current challenges and best-practice protocols for microbiome analysis', *Briefings in Bioinformatics*, 22(1), pp. 178–193. Disponible sur: <https://doi.org/10.1093/bib/bbz155>.
- Bhattacharjya, R. et al. (2023) 'Diatom-assisted aquaculture: Paving the way towards sustainable economy', *Reviews in Aquaculture*, p. raq.12848. Disponible sur: <https://doi.org/10.1111/raq.12848>.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. Disponible sur: <https://doi.org/10.1093/bioinformatics/btu170>.
- Bolyen, E. et al. (2019) 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*, 37(8), pp. 852–857. Disponible sur: <https://doi.org/10.1038/s41587-019-0209-9>.
- Bowler, C. et al. (2008) 'The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes', *Nature*, 456(7219), pp. 239–244. Disponible sur: <https://doi.org/10.1038/nature07410>.
- Bowman, J.P. (2020) 'Out From the Shadows – Resolution of the Taxonomy of the Family Cryomorphaceae', *Frontiers in Microbiology*, 11, p. 795. Disponible sur: <https://doi.org/10.3389/fmicb.2020.00795>.
- Brannan, C.I. et al. (1990) 'The product of the H19 gene may function as an RNA.', *Molecular and Cellular Biology*, 10(1), pp. 28–36. Disponible sur: <https://doi.org/10.1128/MCB.10.1.28>.
- Britten, R.J. and Davidson, E.H. (1969) 'Gene Regulation for Higher Cells: A Theory: New facts regarding the organization of the genome provide clues to the nature of gene regulation.', *Science*, 165(3891), pp. 349–357. Disponible sur: <https://doi.org/10.1126/science.165.3891.349>.

- Bruning, K. (1991) 'Infection of the diatom *Asterionella* by a chytrid. I. Effects of light on reproduction and infectivity of the parasite', *Journal of Plankton Research*, 13(1), pp. 103–117. Disponible sur: <https://doi.org/10.1093/plankt/13.1.103>.
- Brylka, K. et al. (2023) 'Uncertainties surrounding the oldest fossil record of diatoms', *Scientific Reports*, 13(1), p. 8047. Disponible sur: <https://doi.org/10.1038/s41598-023-35078-8>.
- Buhmann, M.T., Day, J.G. and Kroth, P.G. (2013) 'Post-cryopreservation viability of the benthic freshwater diatom *Planothidium frequentissimum* depends on light levels', *Cryobiology*, 67(1), pp. 23–29. Disponible sur: <https://doi.org/10.1016/j.cryobiol.2013.04.005>.
- Cadoret, J.-P. and Bernard, O. (2008) 'La production de biocarburant lipidique avec des microalgues : promesses et défis', *Journal de la Société de Biologie*, 202(3), pp. 201–211. Disponible sur: <https://doi.org/10.1051/jbio:2008022>.
- Centers for Disease Control and Prevention (U.S.) (2019) Antibiotic resistance threats in the United States, 2019. Centers for Disease Control and Prevention (U.S.). Disponible sur: <https://doi.org/10.15620/cdc:82532>.
- Chakravorty, S. et al. (2007) 'A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria', *Journal of Microbiological Methods*, 69(2), pp. 330–339. Disponible sur: <https://doi.org/10.1016/j.mimet.2007.02.005>.
- Chen, L. et al. (2022) 'Short- and long-read metagenomics expand individualized structural variations in gut microbiomes', *Nature Communications*, 13(1), p. 3175. Disponible sur: <https://doi.org/10.1038/s41467-022-30857-9>.
- Chen, Y. et al. (2022) 'Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases', *The Clinical Respiratory Journal*, 16(10), pp. 646–656. Disponible sur: <https://doi.org/10.1111/crj.13538>.
- Chen, Z. and He, X. (2021) 'Application of third-generation sequencing in cancer research', *Medical Review*, 1(2), pp. 150–171. Disponible sur: <https://doi.org/10.1515/mr-2021-0013>.
- Chepurnov, V.A. et al. (2004) 'Experimental Studies on Sexual Reproduction in Diatoms', in *International Review of Cytology*. Elsevier, pp. 91–154. Disponible sur: [https://doi.org/10.1016/S0074-7696\(04\)37003-8](https://doi.org/10.1016/S0074-7696(04)37003-8).
- Chepurnov, V.A. and Mann, D.G. (1997) 'Variation in the sexual behaviour of natural clones of *Achnanthes longipes* (Bacillariophyta)', *European Journal of Phycology*, 32(2), pp. 147–154.
- Clemett, D. and Markham, A. (2000) 'Linezolid', *Drugs*, 59, pp. 815–827.
- Cockerill, F.R. (1999) 'Genetic Methods for Assessing Antimicrobial Resistance', *Antimicrobial Agents and Chemotherapy*, 43(2), pp. 199–212. Disponible sur: <https://doi.org/10.1128/AAC.43.2.199>.
- Cohen, R. (2004) 'Les antibiotiques, c'est pas automatique', *enfances & PSY*, (1), pp. 24–30.

- Compeau, P. and Pevzner, P. (2015) *Bioinformatics algorithms: an active learning approach*. 2nd Edition. La Jolla, CA: Active Learning Publishers.
- Constantinides, B., Hunt, M. and Crook, D.W. (2023) Hostile: accurate host decontamination of microbial sequences. preprint. *Bioinformatics*. Disponible sur: <https://doi.org/10.1101/2023.07.04.547735>.
- Dąbek, P. et al. (2022) 'The blue diatom *Haslea ostrearia* from the Indian Ocean coast of South Africa, with comparative analysis of *Haslea* organellar genomes', *Frontiers in Marine Science*, 9, p. 950716.
- Dahm, R. (2005) 'Friedrich Miescher and the discovery of DNA', *Developmental Biology*, 278(2), pp. 274–288. Disponible sur: <https://doi.org/10.1016/j.ydbio.2004.11.028>.
- D'Alelio, D. et al. (2009) 'Sexual and vegetative phases in the planktonic diatom *Pseudonitzschia multistriata*', *Harmful Algae*, 8(2), pp. 225–232. Disponible sur: <https://doi.org/10.1016/j.hal.2008.05.004>.
- Danecek, P. et al. (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2), p. giab008. Disponible sur: <https://doi.org/10.1093/gigascience/giab008>.
- Davies, J. (1995) 'Vicious circles: looking back on resistance plasmids', *Genetics*, 139(4), p. 1465.
- Davies, J., Farrant, W.N. and Uttley, A.C. (1970) 'Antibiotic resistance of *Shigella sonnei*', *The Lancet*, 296(7684), pp. 1157–1159.
- De Queiroz, K. (2007) 'Species Concepts and Species Delimitation', *Systematic Biology*, 56(6), pp. 879–886. Disponible sur: <https://doi.org/10.1080/10635150701701083>.
- Dekker, C. (2007) 'Solid-state nanopores', *Nature nanotechnology*, 2(4), pp. 209–215.
- Deschamps, S., Llaca, V. and May, G.D. (2012) 'Genotyping-by-Sequencing in Plants', *Biology*, 1(3), pp. 460–483. Disponible sur: <https://doi.org/10.3390/biology1030460>.
- Desclés, J. (2007) *Phaeodactylum tricornutum: un nouveau modèle de diatomée pour l'étude du rôle du silicium dans le monde vivant?* Paris 6.
- Ding, C. and He, J. (2010) 'Effect of antibiotics in the environment on microbial populations', *Applied microbiology and biotechnology*, 87, pp. 925–941.
- Dobin, A. et al. (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. Disponible sur: <https://doi.org/10.1093/bioinformatics/bts635>.
- Donkor, E. (2013) 'Sequencing of Bacterial Genomes: Principles and Insights into Pathogenesis and Development of Antibiotics', *Genes*, 4(4), pp. 556–572. Disponible sur: <https://doi.org/10.3390/genes4040556>.
- Dorrell, R.G. et al. (2017) 'Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome', *elife*, 6, p. e23717.
- Dorrell, R.G. and Bowler, C. (2017) 'Secondary plastids of stramenopiles', in *Advances in botanical research*. Elsevier, pp. 57–103.

- Douglas, A.E. and Werren, J.H. (2016) 'Holes in the hologenome: why host-microbe symbioses are not holobionts', *MBio*, 7(2), p. 10.1128/mbio.02099-15.
- Dubos, R.J. (1937) 'The autolytic system of pneumococci', *The Journal of Experimental Medicine*, 65(6), p. 873.
- Echeverria, P. et al. (1978) 'Antimicrobial resistance and enterotoxin production among isolates of *Escherichia coli* in the Far East', *The Lancet*, 312(8090), pp. 589–592.
- Eickbush, T.H. (2001) 'Retrotransposons', in *Encyclopedia of Genetics*. Elsevier, pp. 1699–1701. Disponible sur: <https://doi.org/10.1006/rwgn.2001.1111>.
- Ellegaard, M. et al. (2016) 'The fascinating diatom frustule—can it play a role for attenuation of UV radiation?', *Journal of Applied Phycology*, 28, pp. 3295–3306.
- Enne, V.I. and Bennett, P.M. (2010) 'Methods to Determine Antibiotic Resistance Gene Silencing', in S.H. Gillespie and T.D. McHugh (eds) *Antibiotic Resistance Protocols*. Totowa, NJ: Humana Press (Methods in Molecular Biology), pp. 29–44. Disponible sur: https://doi.org/10.1007/978-1-60327-279-7_3.
- Enright, M.C. et al. (2002) 'The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA)', *Proceedings of the National Academy of Sciences*, 99(11), pp. 7687–7692. Disponible sur: <https://doi.org/10.1073/pnas.122108599>.
- Ewels, P. et al. (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19), pp. 3047–3048. Disponible sur: <https://doi.org/10.1093/bioinformatics/btw354>.
- Falaise, C. et al. (2019) 'Complex relationships between the blue pigment marennine and marine bacteria of the genus *Vibrio*', *Marine drugs*, 17(3), p. 160.
- Falciatore, A. et al. (2020) 'Diatom Molecular Research Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity', *The Plant Cell*, 32(3), pp. 547–572. Disponible sur: <https://doi.org/10.1105/tpc.19.00158>.
- Ferrante et al. (2019) 'Exploring Molecular Signs of Sex in the Marine Diatom *Skeletonema marinoi*', *Genes*, 10(7), p. 494. Disponible sur: <https://doi.org/10.3390/genes10070494>.
- Fleming, A., Sheehan, J. and JM, A. (1928) 'The History of Penicillin'.
- Frost, L.S. (1992) 'Bacterial conjugation: everybody's doin' it', *Canadian Journal of Microbiology*, 38(11), pp. 1091–1096. Disponible sur: <https://doi.org/10.1139/m92-179>.
- Fry, B. and Wainright, S.C. (1991) 'Diatom sources of 13 C-rich carbon in marine food webs', *Marine Ecology Progress Series*, pp. 149–157.
- Fu, W. et al. (2022) 'Diatom morphology and adaptation: Current progress and potentials for sustainable development', *Sustainable Horizons*, 2, p. 100015. Disponible sur: <https://doi.org/10.1016/j.horiz.2022.100015>.

- Furusawa, G. et al. (2003) 'Algicidal activity and gliding motility of *Saprospira* sp. SS98-5', *Canadian Journal of Microbiology*, 49(2), pp. 92–100. Disponible sur: <https://doi.org/10.1139/w03-017>.
- Gabed, N. et al. (2022) 'What Was Old Is New Again: The Pennate Diatom *Haslea ostrearia* (Gaillon) Simonsen in the Multi-Omic Age', *Marine Drugs*, 20(4), p. 234. Disponible sur: <https://doi.org/10.3390/md20040234>.
- Gabory, A. et al. (2009) 'H19 acts as a trans regulator of the imprinted gene network controlling growth in mice', *Development*, 136(20), pp. 3413–3421. Disponible sur: <https://doi.org/10.1242/dev.036061>.
- Gaillon, M. (1820) 'Des huitres vertes, et des causes de cette coloration.', *Bulletin des Sciences de la Société Philomathique de Paris 1820.*, pp. 129-130. Disponible sur: <https://diatombase.org/aphia.php?p=sourcedetails&id=263822> (Accessed: 11 August 2023).
- Gao, X., Bowler, C. and Kazamia, E. (2021) 'Iron metabolism strategies in diatoms', *Journal of Experimental Botany*. Edited by J. Balk, 72(6), pp. 2165–2180. Disponible sur: <https://doi.org/10.1093/jxb/eraa575>.
- Gastineau, R. et al. (2012) 'Biological Activities of Purified Marennine, the Blue Pigment Responsible for the Greening of Oysters', *Journal of Agricultural and Food Chemistry*, 60(14), pp. 3599–3605. Disponible sur: <https://doi.org/10.1021/jf205004x>.
- Ghenghesh, K.S. et al. (2013) 'Antimicrobial resistance in Libya: 1970-2011', *Libyan Journal of Medicine*, 8(1).
- Goffeau, A. et al. (1996) 'Life with 6000 genes', *Science*, 274(5287), pp. 546–567.
- Gordon, R. and Drum, R.W. (1994) 'Diatom Morphogenesis', *Mechanical Engineering of the Cytoskeleton in Developmental Biology*, 150, p. 421.
- Guo, L. et al. (2015) 'Comparison of potential diatom "barcode" genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta', *International Journal of Systematic and Evolutionary Microbiology*, 65(Pt_4), pp. 1369–1380. Disponible sur: <https://doi.org/10.1099/ijs.0.000076>.
- Gupta, S.K. et al. (2014) 'ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes', *Antimicrobial agents and chemotherapy*, 58(1), pp. 212–220.
- Gurevich, A. et al. (2013) 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics*, 29(8), pp. 1072–1075. Disponible sur: <https://doi.org/10.1093/bioinformatics/btt086>.

- Haas, B.J. et al. (2013) 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature Protocols*, 8(8), pp. 1494–1512. Disponible sur: <https://doi.org/10.1038/nprot.2013.084>.
- Hadziavdic, K. et al. (2014) 'Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers', *PLoS ONE*. Edited by C.R. Voolstra, 9(2), p. e87624. Disponible sur: <https://doi.org/10.1371/journal.pone.0087624>.
- Hajnsdorf, E. and Kaberdin, V.R. (2018) 'RNA polyadenylation and its consequences in prokaryotes', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1762), p. 20180166. Disponible sur: <https://doi.org/10.1098/rstb.2018.0166>.
- Hamm, C.E. et al. (2003) 'Architecture and material properties of diatom shells provide effective mechanical protection', *Nature*, 421(6925), pp. 841–843. Disponible sur: <https://doi.org/10.1038/nature01416>.
- Hamsher, S.E. et al. (2019) 'Extensive chloroplast genome rearrangement amongst three closely related *Halamphora* spp. (Bacillariophyceae), and evidence for rapid evolution as compared to land plants', *PLOS ONE*. Edited by B. Heinze, 14(7), p. e0217824. Disponible sur: <https://doi.org/10.1371/journal.pone.0217824>.
- Harbarth, S. et al. (2015) Antimicrobial resistance: one world, one fight! *BioMed Central*.
- HAS (2014) Principes généraux et conseils de prescription des antibiotiques en premier recours. HAS. Disponible sur: https://www.has-sante.fr/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_1725360.
- Helliwell, K.E., Shibl, A.A. and Amin, S.A. (2022) 'The Diatom Microbiome: New Perspectives for Diatom-Bacteria Symbioses', in A. Falciatore and T. Mock (eds) *The Molecular Life of Diatoms*. Cham: Springer International Publishing, pp. 679–712. Disponible sur: https://doi.org/10.1007/978-3-030-92499-7_23.
- Hildebrand, M. and Lerch, S.J.L. (2015) 'Diatom silica biomineralization: Parallel development of approaches and understanding', *Seminars in Cell & Developmental Biology*, 46, pp. 27–35. Disponible sur: <https://doi.org/10.1016/j.semcdb.2015.06.007>.
- Holmes, R.W. (1985) 'The morphology of diatoms epizoic on cetaceans and their transfer from *Cocconeis* to two new genera, *Bennettella* and *Epipellis*', *British Phycological Journal*, 20(1), pp. 43–57.
- Hou, F. et al. (2020) 'Co-occurrence patterns and assembly processes of microeukaryotic communities in an early-spring diatom bloom', *Science of The Total Environment*, 711, p. 134624. Disponible sur: <https://doi.org/10.1016/j.scitotenv.2019.134624>.
- Hu, J. et al. (2022) 'Differences in diversity and community assembly processes between planktonic and benthic diatoms in the upper reach of the Jinsha River, China', *Hydrobiologia*, 849(7), pp. 1577–1591. Disponible sur: <https://doi.org/10.1007/s10750-022-04801-3>.

- Iglesias-Rodríguez, M.D. et al. (2002) 'Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids', *Global Biogeochemical Cycles*, 16(4), pp. 47-1-47-20.
- Jaillard, M. et al. (2020) 'Interpreting k-mer-based signatures for antibiotic resistance prediction', *GigaScience*, 9(10), p. giaa110. Disponible sur: <https://doi.org/10.1093/gigascience/giaa110>.
- Jauffrais, T. et al. (2017) 'Effect of bacteria on growth and biochemical composition of two benthic diatoms *Halamphora coffeaeformis* and *Entomoneis paludosa*', *Journal of Experimental Marine Biology and Ecology*, 495, pp. 65–74. Disponible sur: <https://doi.org/10.1016/j.jembe.2017.06.004>.
- Jevons, M.P. (1961) "Celbenin"-resistant staphylococci', *British medical journal*, 1(5219), p. 124.
- JGI (2011) Summary statistics for the *Pseudo-nitzschia* multiseres v1.0 release. Disponible sur: <https://mycocosm.jgi.doe.gov/Psemu1/Psemu1.info.html> (Accessed: 15 September 2023).
- Jiang, C. et al. (2023) 'iGDP : An integrated genome decontamination pipeline for wild ciliated microeukaryotes', *Molecular Ecology Resources*, 23(5), pp. 1182–1193. Disponible sur: <https://doi.org/10.1111/1755-0998.13782>.
- Johansson, O.N. et al. (2019) 'Friends With Benefits: Exploring the Phycosphere of the Marine Diatom *Skeletonema marinoi*', *Frontiers in Microbiology*, 10, p. 1828. Disponible sur: <https://doi.org/10.3389/fmicb.2019.01828>.
- Jones, A.K., Rhodes, M.E. and Evans, S.C. (1973) 'The use of antibiotics to obtain axenic cultures of algae', *British Phycological Journal*, 8(2), pp. 185–196. Disponible sur: <https://doi.org/10.1080/00071617300650211>.
- Jurka, J. et al. (2005) 'Repbase Update, a database of eukaryotic repetitive elements', *Cytogenetic and Genome Research*, 110(1–4), pp. 462–467. Disponible sur: <https://doi.org/10.1159/000084979>.
- Karas, B.J. et al. (2015) 'Designer diatom episomes delivered by bacterial conjugation', *Nature communications*, 6(1), p. 6925.
- Kermarrec, L. et al. (2011) 'A preliminary phylogenetic analysis of the Cymbellales based on 18S rDNA gene sequencing', *Diatom Research*, 26(3), pp. 305–315. Disponible sur: <https://doi.org/10.1080/0269249X.2011.633255>.
- Khan, S. et al. (2020) 'Effects of epiphytic diatom *Licmophora paradoxa* on photosynthesis, malondialdehyde content and antioxidant enzymes activities of *Pyropia yezoensis*'.
- Kim, B.-H. et al. (2005) 'Antialgal interactions of biological control agents on cyanobacterium and diatom blooms in vitro', *Korean J Ecol Environ*, 38, pp. 494–502.

- Klindworth, A. et al. (2014) 'Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing', *Marine Genomics*, 18, pp. 185–192. Disponible sur: <https://doi.org/10.1016/j.margen.2014.08.007>.
- Koedooder, C. et al. (2019) 'Diatom-Bacteria Interactions Modulate the Composition and Productivity of Benthic Diatom Biofilms', *Frontiers in Microbiology*, 10, p. 1255. Disponible sur: <https://doi.org/10.3389/fmicb.2019.01255>.
- Kolmogorov, M. et al. (2019) 'Assembly of long, error-prone reads using repeat graphs', *Nature biotechnology*, 37(5), pp. 540–546.
- Kolmogorov, M. et al. (2020) 'metaFlye: scalable long-read metagenome assembly using repeat graphs', *Nature Methods*, 17(11), pp. 1103–1110. Disponible sur: <https://doi.org/10.1038/s41592-020-00971-x>.
- Kooistra, W.H. et al. (2007) 'The origin and evolution of the diatoms: their adaptation to a planktonic existence', *Evolution of primary producers in the sea*, pp. 207–249.
- Kooistra, W.H. and Medlin, L.K. (1996) 'Evolution of the diatoms (Bacillariophyta): IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record', *Molecular phylogenetics and evolution*, 6(3), pp. 391–407.
- Koren, S. et al. (2017) 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome research*, 27(5), pp. 722–736.
- Koziaeva, V.V. et al. (2023) 'Magnetospirillum sulfuroxidans sp. nov., capable of sulfur-dependent lithoautotrophy and a taxonomic reevaluation of the order Rhodospirillales', *Systematic and Applied Microbiology*, 46(3), p. 126406. Disponible sur: <https://doi.org/10.1016/j.syapm.2023.126406>.
- Kudoh, S. and Takahashi, M. (1989) 'Physico-chemical control of the growth of a diatom, *Asterionella formosa* Hass., in a shallow eutrophic lake', *Journal of Plankton Research*, 11(5), pp. 1001–1019. Disponible sur: <https://doi.org/10.1093/plankt/11.5.1001>.
- Kützing, F.T. (1844) *Die kieselschaligen Bacillarien oder Diatomeen: (Mit 30 vom Verfasser gravirten Tafeln)*. W. Köhne.
- Kuznetsov, D. et al. (2023) 'OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity', *Nucleic Acids Research*, 51(D1), pp. D445–D451. Disponible sur: <https://doi.org/10.1093/nar/gkac998>.
- Lataretu, M. et al. (2023) Targeted decontamination of sequencing data with CLEAN. preprint. *Bioinformatics*. Disponible sur: <https://doi.org/10.1101/2023.08.05.552089>.
- Latorre-Pérez, A. et al. (2020) 'Assembly methods for nanopore-based metagenomic sequencing: a comparative study', *Scientific Reports*, 10(1), p. 13588. Disponible sur: <https://doi.org/10.1038/s41598-020-70491-3>.

- Latour, J.C.-T. et al. (2023) 'Effect of marennine on the rearing medium and microbiota of *Mytilus edulis* larvae and its protective effect after exposure to the pathogenic *Vibrio splendidus*', *Journal of Applied Aquaculture*, pp. 1–21. Disponible sur: <https://doi.org/10.1080/10454438.2023.2236096>.
- Lebeau, T. and Robert, J.-M. (2003) 'Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products', *Applied microbiology and biotechnology*, 60, pp. 624–632.
- Lépinay, A. et al. (2016) 'Bacterial community structure of the marine diatom *Haslea ostrearia*', *Algal Research*, 16, pp. 418–426. Disponible sur: <https://doi.org/10.1016/j.algal.2016.04.011>.
- Lépinay, A. et al. (2018) 'First insight on interactions between bacteria and the marine diatom *Haslea ostrearia*: Algal growth and metabolomic fingerprinting', *Algal Research*, 31, pp. 395–405. Disponible sur: <https://doi.org/10.1016/j.algal.2018.02.023>.
- Letáková, M., Fránková, M. and Pouličková, A. (2018) 'Ecology and Applications of Freshwater Epiphytic Diatoms — Review', *Cryptogamie, Algologie*, 39(1), pp. 3–22. Disponible sur: <https://doi.org/10.7872/crya/v39.iss1.2018.3>.
- Lewitus, E. et al. (2018) 'Clade-specific diversification dynamics of marine diatoms since the Jurassic', *Nature Ecology & Evolution*, 2(11), pp. 1715–1723. Disponible sur: <https://doi.org/10.1038/s41559-018-0691-3>.
- Li, H. (2018) 'Minimap2: pairwise alignment for nucleotide sequences', *Bioinformatics*. Edited by I. Birol, 34(18), pp. 3094–3100. Disponible sur: <https://doi.org/10.1093/bioinformatics/bty191>.
- Lian, J. et al. (2021) 'Different co-occurring bacteria enhance or decrease the growth of the microalga *Nannochloropsis* sp. CCAP211/78', *Microbial Biotechnology*, 14(3), pp. 1159–1170. Disponible sur: <https://doi.org/10.1111/1751-7915.13784>.
- Ligon, B.L. (2004) 'Penicillin: its discovery and early development', in *Seminars in pediatric infectious diseases*. Elsevier, pp. 52–57.
- Lina, G. et al. (1999) 'Distribution of genes encoding resistance to macrolides, lincosamides, and streptogramins among staphylococci', *Antimicrobial agents and chemotherapy*, 43(5), pp. 1062–1066.
- Liu, Y., Wang, X. and Sun, J. (2022) 'Bacterial Transformation and Processing of Diatom-Derived Organic Matter: A Case Study for *Skeletonema dohrnii*', *Frontiers in Microbiology*, 13, p. 840564. Disponible sur: <https://doi.org/10.3389/fmicb.2022.840564>.
- Liu, S., Xu, Q. and Chen, N. (2023) 'Expansion of photoreception-related gene families may drive ecological adaptation of the dominant diatom species *Skeletonema marinoi*', *Science of The Total Environment*, 897, p. 165384. Disponible sur: <https://doi.org/10.1016/j.scitotenv.2023.165384>.

- Macdonald, J.D. (1869) 'I.—On the structure of the Diatomaceous frustule, and its genetic cycle', *Journal of Natural History*, 3(13), pp. 1–8.
- Maeda, Y. et al. (2022) 'Chromosome-Scale Genome Assembly of the Marine Oleaginous Diatom *Fistulifera solaris*', *Marine Biotechnology*, 24(4), pp. 788–800. Disponible sur: <https://doi.org/10.1007/s10126-022-10147-7>.
- Malviya, S. et al. (2016) 'Insights into global diatom distribution and diversity in the world's ocean', *Proceedings of the National Academy of Sciences*, 113(11). Disponible sur: <https://doi.org/10.1073/pnas.1509523113>.
- Mangalea, M.R. et al. (2023) 'Assembly and Annotation of Viral Metagenomes from Short-Read Sequencing Data', in *Metagenomic Data Analysis*. Springer, pp. 317–337.
- Mann, D.G. (1978) *Studies in the Nitzschiaceae (Bacillariophyta)*.
- Mann, D.G. and Chepurnov, V.A. (2004) 'What have the Romans ever done for us? The past and future contribution of culture studies to diatom systematics', *Nova Hedwigia*, pp. 237–291.
- Mann, D.G. and Droop, S.J.M. (1996) 'Biodiversity, biogeography and conservation of diatoms', in *Biogeography of Freshwater Algae: Proceedings of the Workshop on Biogeography of Freshwater Algae, Held During the Fifth International Phycological Congress, Qingdao, China, June 1994*. Springer, pp. 19–32.
- Mann, D.G. and Vanormelingen, P. (2013) 'An inordinate fondness? The number, distributions, and origins of diatom species', *Journal of eukaryotic microbiology*, 60(4), pp. 414–420.
- Mann, S. and Chen, Y.-P.P. (2010) 'Bacterial genomic G+C composition-eliciting environmental adaptation', *Genomics*, 95(1), pp. 7–15. Disponible sur: <https://doi.org/10.1016/j.ygeno.2009.09.002>.
- Manni, M. et al. (2021) 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes', *Molecular Biology and Evolution*. Edited by J. Kelley, 38(10), pp. 4647–4654. Disponible sur: <https://doi.org/10.1093/molbev/msab199>.
- Marçais, G. and Kingsford, C. (2011) 'A fast, lock-free approach for efficient parallel counting of occurrences of k -mers', *Bioinformatics*, 27(6), pp. 764–770. Disponible sur: <https://doi.org/10.1093/bioinformatics/btr011>.
- Marshall, S.M. and Orr, A.P. (1928) 'The photosynthesis of diatom cultures in the sea', *Journal of the Marine Biological Association of the United Kingdom*, 15(1), pp. 321–360.
- Martín, R. et al. (2014) 'The role of metagenomics in understanding the human microbiome in health and disease', *Virulence*, 5(3), pp. 413–423. Disponible sur: <https://doi.org/10.4161/viru.27864>.

- Martin, W. and Kowallik, K.V. (1999) 'Annotated English translation of Mereschkowsky's 1905 paper "Über Natur und Ursprung der Chromatophoren im Pflanzenreiche"', *European Journal of Phycology*, 34(3), pp. 287–295.
- Martino, A.D. et al. (2007) 'Genetic and phenotypic characterization of *Phaeodactylum tricorutum* (Bacillariophyceae) accessions 1', *Journal of Phycology*, 43(5), pp. 992–1009. Disponible sur: <https://doi.org/10.1111/j.1529-8817.2007.00384.x>.
- Maunsell, T.J. et al. (2021) A study of the correlation between phenotypic antimicrobial susceptibility testing methods and the associated genotypes determined by whole genome sequencing for a collection of *Escherichia coli* of bovine origin. preprint. *Microbiology*. Disponible sur: <https://doi.org/10.1101/2021.07.16.452668>.
- Maxam, A.M. and Gilbert, W. (1980) '[57] Sequencing end-labeled DNA with base-specific chemical cleavages', in *Methods in Enzymology*. Elsevier, pp. 499–560. Available at: [https://doi.org/10.1016/S0076-6879\(80\)65059-9](https://doi.org/10.1016/S0076-6879(80)65059-9).
- Mayr, E. (1999) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.
- McArthur, A.G. et al. (2013) 'The Comprehensive Antibiotic Resistance Database', *Antimicrobial Agents and Chemotherapy*, 57(7), pp. 3348–3357. Disponible sur: <https://doi.org/10.1128/AAC.00419-13>.
- McArthur, A.G. and Wright, G.D. (2015) 'Bioinformatics of antimicrobial resistance in the age of molecular epidemiology', *Current Opinion in Microbiology*, 27, pp. 45–50. Disponible sur: <https://doi.org/10.1016/j.mib.2015.07.004>.
- McFadden, G.I. (2001) 'PRIMARY AND SECONDARY ENDOSYMBIOSIS AND THE ORIGIN OF PLASTIDS', *Journal of Phycology*, 37(6), pp. 951–959. Disponible sur: <https://doi.org/10.1046/j.1529-8817.2001.01126.x>.
- McNulty, C.A.M. et al. (2007) 'Don't wear me out--the public's knowledge of and attitudes to antibiotic use', *Journal of Antimicrobial Chemotherapy*, 59(4), pp. 727–738. Disponible sur: <https://doi.org/10.1093/jac/dkl558>.
- Medlin, L.K. et al. (1996) 'Evolution of the diatoms (Bacillariophyta). II. Nuclear-encoded small-subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms', *Molecular Biology and Evolution*, 13(1), pp. 67–75. Disponible sur: <https://doi.org/10.1093/oxfordjournals.molbev.a025571>.
- Medlin, L.K. et al. (2008) 'MOLECULAR EVIDENCE CONFIRMS SISTER RELATIONSHIP OF ARDISSONEA, CLIMACOSPHENIA, AND TOXARIUM WITHIN THE BIPOLAR CENTRIC DIATOMS (BACILLARIOPHYTA, MEDIOPHYCEAE), AND CLADISTIC ANALYSES CONFIRM THAT EXTREMELY ELONGATED SHAPE HAS ARISEN TWICE IN THE DIATOMS 1', *Journal of Phycology*, 44(5), pp. 1340–1348.

- Medlin, L.K. (2016) 'Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence', *Phycologia*, 55(1), pp. 79–103. Disponible sur: <https://doi.org/10.2216/15-105.1>.
- Medlin, L.K. (2018) 'Mini review: Diatom species as seen through a molecular window', *Brazilian Journal of Botany*, 41(2), pp. 457–469. Disponible sur: <https://doi.org/10.1007/s40415-018-0444-1>.
- Medlin, L.K. and Desdevises, Y. (2020) 'Phylogenetic reconstruction of diatoms using a seven-gene dataset, multiple outgroups, and morphological data for a total evidence approach', *Phycologia*, 59(5), pp. 422–436. Disponible sur: <https://doi.org/10.1080/00318884.2020.1795962>.
- Medlin, L.K. and Kaczmarek, I. (2004) 'Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision', *Phycologia*, 43(3), pp. 245–270. Disponible sur: <https://doi.org/10.2216/i0031-8884-43-3-245.1>.
- Mereschkowsky, C. (1905) 'Über natur und ursprung der chromatophoren im pflanzenreiche', *Biol Centralbl*, 25, p. 593.
- Méresse, S. et al. (2023) 'Haslea ostrearia Pigment Marennine Affects Key Actors of Neuroinflammation and Decreases Cell Migration in Murine Neuroglial Cell Model', *International Journal of Molecular Sciences*, 24(6), p. 5388. Disponible sur: <https://doi.org/10.3390/ijms24065388>.
- Miescher, F. (1869) 'Letter i; to wilhelm his; tübingen, february 26th, 1869', *Die histochemischen und physiologischen arbeiten von Friedrich Miescher-aus dem wissenschaftlichen Briefwechsel von F. Miescher*, 1, pp. 33–38.
- Mo, W.Y. et al. (2017) 'Application of veterinary antibiotics in China's aquaculture industry and their potential human health risks', *Environmental Science and Pollution Research*, 24, pp. 8978–8989.
- Mock, T. (2020) 100 Diatom Genomes Project. Disponible sur: <https://jgi.doe.gov/csp-2021-100-diatom-genomes/> (Visité le: 10 November 2023).
- Moeys, S. et al. (2016) 'A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*', *Scientific Reports*, 6(1), p. 19252. Disponible sur: <https://doi.org/10.1038/srep19252>.
- Molisch, H. (1937) 'Effect of one plant on another plant (Allelopathy)', *Forschungen u Forstschr*, 13(34), pp. 407–08.
- Moniz, M.B.J. and Kaczmarek, I. (2010) 'Barcoding of Diatoms: Nuclear Encoded ITS Revisited', *Protist*, 161(1), pp. 7–34. Disponible sur: <https://doi.org/10.1016/j.protis.2009.07.001>.
- Moon, B.Y. et al. (2015) 'Phage-mediated horizontal transfer of a *Staphylococcus aureus* virulence-associated genomic island', *Scientific reports*, 5(1), p. 9784.

- Moore, E.R. et al. (2017) 'Morphological and transcriptomic evidence for ammonium induction of sexual reproduction in *Thalassiosira pseudonana* and other centric diatoms', PLOS ONE. Edited by D.A. Campbell, 12(7), p. e0181098. Disponible sur: <https://doi.org/10.1371/journal.pone.0181098>.
- Moore, J.K. et al. (2001) 'An intermediate complexity marine ecosystem model for the global domain', Deep Sea Research Part II: Topical Studies in Oceanography, 49(1–3), pp. 403–462.
- Moran, N.A. and Sloan, D.B. (2015) 'The hologenome concept: helpful or hollow?', PLoS biology, 13(12), p. e1002311.
- Moreau, J. (1967) 'Recherches préliminaires sur le verdissement en claires: l'évolution de leurs divers pigments liée au complexe pigmentaire de *Navicula ostrearia* Bory', Revue des Travaux de l'Institut des Pêches Maritimes, 314(4), pp. 373–382.
- Mouget, J.-L. et al. (1999) 'Long-term photoacclimation of *Haslea ostrearia* (Bacillariophyta): effect of irradiance on growth rates, pigment content and photosynthesis', European Journal of Phycology, 34(2), pp. 109–115.
- Mouget, J.-L. et al. (2009) 'Light is a key factor in triggering sexual reproduction in the pennate diatom *Haslea ostrearia*', FEMS Microbiology Ecology, 69(2), pp. 194–201.
- Muggli, D., Lecourt, M. and Harrison, P. (1996) 'Effects of iron and nitrogen source on the sinking rate, physiology and metal composition of an oceanic diatom from the subarctic Pacific', Marine Ecology Progress Series, 132, pp. 215–227. Disponible sur: <https://doi.org/10.3354/meps132215>.
- Munyaka, P.M. et al. (2015) 'Antepartum Antibiotic Treatment Increases Offspring Susceptibility to Experimental Colitis: A Role of the Gut Microbiota', PLOS ONE. Edited by M.M. Heimesaat, 10(11), p. e0142536. Disponible sur: <https://doi.org/10.1371/journal.pone.0142536>.
- Murray, C.J.L. et al. (2022) 'Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis', The Lancet, 399(10325), pp. 629–655. Disponible sur: [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- Nakabachi, A. et al. (2006) 'The 160-kilobase genome of the bacterial endosymbiont *Carsonella*', Science, 314(5797), pp. 267–267.
- Nanjappa, D. et al. (2017) 'Diatom flagellar genes and their expression during sexual reproduction in *Leptocylindrus danicus*', BMC Genomics, 18(1), p. 813. Disponible sur: <https://doi.org/10.1186/s12864-017-4210-8>.
- Neefs, J.-M. et al. (1993) 'Compilation of small ribosomal subunit RNA structures', Nucleic Acids Research, 21(13), pp. 3025–3049. Disponible sur: <https://doi.org/10.1093/nar/21.13.3025>.

- Nghiem Xuan, R. et al. (2020) 'Design of an artificial culture medium to optimize *Haslea ostrearia* biomass and marennine production', *Algal Research*, 45, p. 101653. Disponible sur: <https://doi.org/10.1016/j.algal.2019.101653>.
- Nghiem-Xuan, R. (2019) *Optimisation de la culture d'Haslea ostrearia en photobioréacteur*. Nantes.
- Nieri, P. et al. (2023) 'Bioactive Molecules from Marine Diatoms and Their Value for the Nutraceutical Industry', *Nutrients*, 15(2), p. 464. Disponible sur: <https://doi.org/10.3390/nu15020464>.
- Nonoyama, T. et al. (2019) 'Metabolic Innovations Underpinning the Origin and Diversification of the Diatom Chloroplast', *Biomolecules*, 9(8), p. 322. Disponible sur: <https://doi.org/10.3390/biom9080322>.
- Norman, A., Hansen, L.H. and Sørensen, S.J. (2009) 'Conjugative plasmids: vessels of the communal gene pool', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1527), pp. 2275–2289. Disponible sur: <https://doi.org/10.1098/rstb.2009.0037>.
- Nurk, S. et al. (2022) 'The complete sequence of a human genome', *Science*, 376(6588), pp. 44–53. Disponible sur: <https://doi.org/10.1126/science.abj6987>.
- Orozco-Arias, S. et al. (2018) 'Inpactor, Integrated and Parallel Analyzer and Classifier of LTR Retrotransposons and Its Application for Pineapple LTR Retrotransposons Diversity and Dynamics', *Biology*, 7(2), p. 32. Disponible sur: <https://doi.org/10.3390/biology7020032>.
- Otte, A. et al. (2023) 'The diatom *Fragilariopsis cylindrus*: A model alga to understand cold-adapted life', *Journal of Phycology*, 59(2), pp. 301–306. Disponible sur: <https://doi.org/10.1111/jpy.13325>.
- Oudot-Le Secq, M.-P. et al. (2007) 'Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage', *Molecular Genetics and Genomics*, 277(4), pp. 427–439. Disponible sur: <https://doi.org/10.1007/s00438-006-0199-4>.
- Palmer, A.C. and Kishony, R. (2013) 'Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance', *Nature Reviews Genetics*, 14(4), pp. 243–248. Disponible sur: <https://doi.org/10.1038/nrg3351>.
- Poulin, M., Méléder, V. and Mouget, J.-L. (2019) 'Typification of the first recognized blue pigmented diatom, *Haslea ostrearia* (Bacillariophyceae)', *Plant Ecology and Evolution*, 152(2), pp. 402–408. Disponible sur: <https://doi.org/10.5091/plecevo.2019.1622>.
- Pouvreau, J.-B. et al. (2006) 'Purification of the blue-green pigment "marennine" from the marine tychopelagic diatom *Haslea ostrearia* (Gaillon/Bory) Simonsen', *Journal of Applied Phycology*, 18(6), pp. 769–781. Disponible sur: <https://doi.org/10.1007/s10811-006-9088-9>.

- Prasetya, F.S. et al. (2020) 'Effect of the blue pigment produced by the tropical diatom *Haslea nusantara* on marine organisms from different trophic levels and its bioactivity', *Aquaculture Reports*, 17, p. 100389. Disponible sur: <https://doi.org/10.1016/j.aqrep.2020.100389>.
- Prazukin, A.V. et al. (2022) 'Vertical distribution of epiphytic diatoms in relation to the eelgrass *Zostera noltii* canopy biomass and height', *Aquatic Botany*, 176, p. 103466. Disponible sur: <https://doi.org/10.1016/j.aquabot.2021.103466>.
- Prihoda, J. et al. (2012) 'Chloroplast-mitochondria cross-talk in diatoms', *Journal of experimental botany*, 63(4), pp. 1543–1557.
- Purushothaman, A. et al. (2021) 'First report of marine epizoic diatom, *Protoraphis atlantica* (Protoraphidaceae) on calanoid copepods along the southeastern Arabian Sea', *Symbiosis*, 84(2), pp. 131–140. Disponible sur: <https://doi.org/10.1007/s13199-021-00772-6>.
- Quast, C. et al. (2012) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research*, 41(D1), pp. D590–D596. Disponible sur: <https://doi.org/10.1093/nar/gks1219>.
- Rabosky, D.L. and Sorhannus, U. (2009) 'Diversity dynamics of marine planktonic diatoms across the Cenozoic', *Nature*, 457(7226), pp. 183–186. Disponible sur: <https://doi.org/10.1038/nature07435>.
- Ragueneau, O. et al. (2006) 'Role of diatoms in silicon cycling and coastal marine food webs', *The silicon cycle: human perturbations and impacts on aquatic systems*, 66, pp. 163–195.
- Rammelkamp, C.H. and Weinstein, L. (1942) 'Toxic effects of tyrothricin, gramicidin and tyrocidine', *The Journal of Infectious Diseases*, 71(2), pp. 166–173.
- Raven, J.A. and Waite, A.M. (2004) 'The evolution of silicification in diatoms: inescapable sinking and sinking as escape?', *New phytologist*, 162(1), pp. 45–61.
- Retnaningdyah, C., Endang Arisoelaningsih, E. and Samino, S. (2017) 'Use of local Hydromacrophytes as phytoremediation agent in pond to improve irrigation water quality evaluated by Diatom Biotic Indices', *Biodiversitas Journal of Biological Diversity*, 18(4), pp. 1611–1617. Disponible sur: <https://doi.org/10.13057/biodiv/d180439>.
- Rhoads, A. and Au, K.F. (2015) 'PacBio Sequencing and Its Applications', *Genomics, Proteomics & Bioinformatics*, 13(5), pp. 278–289. Disponible sur: <https://doi.org/10.1016/j.gpb.2015.08.002>.
- Rice, P., Bleasby, A. and Ison, J. (2011) *EMBOSS user's guide: practical bioinformatics*. Cambridge ; New York: Cambridge University Press.

- Rivkin, R.B. (1986) 'Radioisotopic method for measuring cell division rates of individual species of diatoms from natural populations', *Applied and environmental microbiology*, 51(4), pp. 769–775.
- Ronaghi, M. et al. (1998) 'PCR-introduced loop structure as primer in DNA sequencing', *Biotechniques*, 25(5), pp. 876–884.
- Rothpletz, A. (1896) 'Ueber die Flysch-Fucoiden und einige andere fossile Algen, sowie über liasische, Diatomeen führende Hornschwämme.', *Zeitschrift der Deutschen Geologischen Gesellschaft*, pp. 854–914.
- Rothpletz, A. (1900) 'Ueber einen neuen jurassischen Hornschwamm und die darin eingeschlossenen Diatomeen.', *Zeitschrift der Deutschen Geologischen Gesellschaft*, pp. 154–160.
- Round, F.E., Crawford, R.M. and Mann, D.G. (1990) *The Diatoms: biology & morphology of the genera*. Cambridge [England]; New York: Cambridge University Press.
- Ruan, J. and Li, H. (2020) 'Fast and accurate long-read assembly with wtdbg2', *Nature Methods*, 17(2), pp. 155–158. Disponible sur: <https://doi.org/10.1038/s41592-019-0669-3>.
- Ruck, E.C. et al. (2016) 'Phylogeny, ecology, morphological evolution, and reclassification of the diatom orders Surirellales and Rhopalodiales', *Molecular Phylogenetics and Evolution*, 103, pp. 155–171. Disponible sur: <https://doi.org/10.1016/j.ympev.2016.07.023>.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proceedings of the national academy of sciences*, 74(12), pp. 5463–5467.
- Sapp, M., Wichels, A. and Gerdt, G. (2007) 'Impacts of Cultivation of Marine Diatoms on the Associated Bacterial Community', *Applied and Environmental Microbiology*, 73(9), pp. 3117–3120. Disponible sur: <https://doi.org/10.1128/AEM.02274-06>.
- Scalco, E. et al. (2014) 'The dynamics of sexual phase in the marine diatom *Pseudo-nitzschia multistriata* (Bacillariophyceae)', *Journal of Phycology*, 50(5), pp. 817–828.
- Schwengers, O. et al. (2020) 'Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores', *Microbial Genomics*, 6(10). Disponible sur: <https://doi.org/10.1099/mgen.0.000398>.
- Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. Disponible sur: <https://doi.org/10.1093/bioinformatics/btu153>.
- Service, R.F. (2006) *The race for the \$1000 genome*. American Association for the Advancement of Science.
- Seveno, J. (2022) *Déterminisme des blooms de *Haslea* en environnement naturel : dynamique et impacts*. Le Mans Université.

- Shao, Q. et al. (2020) 'Succession of bacterioplankton communities over complete Gymnodinium-diatom bloom cycles', *Science of The Total Environment*, 709, p. 135951. Disponible sur: <https://doi.org/10.1016/j.scitotenv.2019.135951>.
- Shibl, A.A. et al. (2020) 'Diatom modulation of select bacteria through use of two unique secondary metabolites', *Proceedings of the National Academy of Sciences*, 117(44), pp. 27445–27455. Disponible sur: <https://doi.org/10.1073/pnas.2012088117>.
- Sigurdson, G.T., Tang, P. and Giusti, M.M. (2017) 'Natural colorants: Food colorants from natural sources', *Annual review of food science and technology*, 8, pp. 261–280.
- Simonsen, R. (1974) *The Diatom Plankton of the Indian Ocean Expedition of RV 'Meteor' 1964-1965*. Gebrüder Borntraeger (Biology, 19).
- Simpson, J.T. et al. (2009) 'ABYSS: a parallel assembler for short read sequence data', *Genome research*, 19(6), pp. 1117–1123.
- Singer, M.F. (1968) '1968 Nobel Laureate in Medicine or Physiology', *Science*, 162(3852), pp. 433–436.
- Slatko, B.E., Gardner, A.F. and Ausubel, F.M. (2018) 'Overview of Next-Generation Sequencing Technologies', *Current Protocols in Molecular Biology*, 122(1). Disponible sur: <https://doi.org/10.1002/cpmb.59>.
- Smith, L.M. et al. (1985) 'The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis', *Nucleic Acids Research*, 13(7), pp. 2399–2412. Disponible sur: <https://doi.org/10.1093/nar/13.7.2399>.
- Sogin, M.L. et al. (2006) 'Microbial diversity in the deep sea and the underexplored "rare biosphere"', *Proceedings of the National Academy of Sciences*, 103(32), pp. 12115–12120. Disponible sur: <https://doi.org/10.1073/pnas.0605127103>.
- Sterling, A.R. et al. (2023) 'Potential interactions between diatoms and bacteria are shaped by trace element gradients in the Southern Ocean', *Frontiers in Marine Science*, 9, p. 876830. Disponible sur: <https://doi.org/10.3389/fmars.2022.876830>.
- Stock, W. et al. (2018) 'Expanding the toolbox for cryopreservation of marine and freshwater diatoms', *Scientific Reports*, 8(1), p. 4279. Disponible sur: <https://doi.org/10.1038/s41598-018-22460-0>.
- Strommenger, B. et al. (2003) 'Multiplex PCR Assay for Simultaneous Detection of Nine Clinically Relevant Antibiotic Resistance Genes in *Staphylococcus aureus*', *Journal of Clinical Microbiology*, 41(9), pp. 4089–4094. Disponible sur: <https://doi.org/10.1128/JCM.41.9.4089-4094.2003>.
- Sueoka, N. (1962) 'ON THE GENETIC BASIS OF VARIATION AND HETEROGENEITY OF DNA BASE COMPOSITION', *Proceedings of the National Academy of Sciences*, 48(4), pp. 582–592. Disponible sur: <https://doi.org/10.1073/pnas.48.4.582>.

- Sultana, S. et al. (2023) 'Vitamin B12 is not shared by all marine prototrophic bacteria with their environment', *The ISME Journal*, 17(6), pp. 836–845. Disponible sur: <https://doi.org/10.1038/s41396-023-01391-3>.
- Sun, J. et al. (2011) 'Effects of changing p CO₂ and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseriata*', *Limnology and Oceanography*, 56(3), pp. 829–840. Disponible sur: <https://doi.org/10.4319/lo.2011.56.3.0829>.
- Sun, Z. et al. (2014) 'Community dynamics of prokaryotic and eukaryotic microbes in an estuary reservoir', *Scientific Reports*, 4(1), p. 6966. Disponible sur: <https://doi.org/10.1038/srep06966>.
- Syama, K., Jothi, J.A.A. and Khanna, N. (2023) 'Automatic disease prediction from human gut metagenomic data using boosting GraphSAGE', *BMC Bioinformatics*, 24(1), p. 126. Disponible sur: <https://doi.org/10.1186/s12859-023-05251-x>.
- Tanniou, A., Turpin, V. and Lebeau, T. (2012) 'Comparison of cryopreservation methods for the long term storage of the marine diatom *Haslea ostrearia* (simonsen)', *Cryobiology*, 65(1), pp. 45–50. Disponible sur: <https://doi.org/10.1016/j.cryobiol.2012.03.011>.
- Teeling, H. et al. (2012) 'Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom', *Science*, 336(6081), pp. 608–611. Disponible sur: <https://doi.org/10.1126/science.1218344>.
- Teuber, M. (2001) 'Veterinary use and antibiotic resistance', *Current Opinion in Microbiology*, 4(5), pp. 493–499. Disponible sur: [https://doi.org/10.1016/S1369-5274\(00\)00241-1](https://doi.org/10.1016/S1369-5274(00)00241-1).
- Theriot, E.C. et al. (2009) 'The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny', *European Journal of Phycology*, 44(3), pp. 277–290. Disponible sur: <https://doi.org/10.1080/09670260902749159>.
- Theriot, E.C. et al. (2010) 'A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research', *Plant Ecology and Evolution*, 143(3), pp. 278–296. Disponible sur: <https://doi.org/10.5091/plecevo.2010.418>.
- Theriot, E.C. et al. (2015) 'Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling', *Molecular Phylogenetics and Evolution*, 89, pp. 28–36. Disponible sur: <https://doi.org/10.1016/j.ympev.2015.03.012>.
- Thines, M. et al. (2015) 'The diatom parasite *Lagenisma coscinodisci* (Lagenismatales, Oomycota) is an early diverging lineage of the Saprolegniomycetes', *Mycological Progress*, 14(9), p. 75. Disponible sur: <https://doi.org/10.1007/s11557-015-1099-y>.
- Tiedje, J.M. et al. (2023) 'Antibiotic resistance genes in food production systems support One Health opinions', *Current Opinion in Environmental Science & Health*, p. 100492.
- Tiffany, M.A. (2011) 'Epizoic and Epiphytic Diatoms', in J. Seckbach and P. Kociolek (eds) *The Diatom World*. Dordrecht: Springer Netherlands (Cellular Origin, Life in Extreme Habitats

- and Astrobiology), pp. 195–209. Disponible sur: https://doi.org/10.1007/978-94-007-1327-7_8.
- Tilstone, G. et al. (2000) 'Diatom dynamics in a coastal ecosystem affected by upwelling: coupling between species succession, circulation and biogeochemical processes', *Marine Ecology Progress Series*, 205, pp. 23–41. Disponible sur: <https://doi.org/10.3354/meps205023>.
- Tirichine, L., Rastogi, A. and Bowler, C. (2017) 'Recent progress in diatom genomics and epigenomics', *Current Opinion in Plant Biology*, 36, pp. 46–55. Available at: <https://doi.org/10.1016/j.pbi.2017.02.001>.
- Touchon, M., De Sousa, J.A.M. and Rocha, E.P. (2017) 'Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer', *Current opinion in microbiology*, 38, pp. 66–73.
- Truong, W.R. et al. (2021) 'The antibiogram: key considerations for its development and utilization', *JAC-Antimicrobial Resistance*, 3(2), p. dlab060. Disponible sur: <https://doi.org/10.1093/jacamr/dlab060>.
- Tuan, N.N. et al. (2014) 'Multiple approaches to characterize the microbial community in a thermophilic anaerobic digester running on swine manure: A case study', *Microbiological Research*, 169(9–10), pp. 717–724. Disponible sur: <https://doi.org/10.1016/j.micres.2014.02.003>.
- Turk Dermastia, T. et al. (2023) 'Evaluation of the rbcL marker for metabarcoding of marine diatoms and inference of population structure of selected genera', *Frontiers in Microbiology*, 14, p. 1071379. Disponible sur: <https://doi.org/10.3389/fmicb.2023.1071379>.
- Vilella, A. (2023) 'Next-Generation-Sequencing'. Disponible sur: https://www.google.com/url?q=https://docs.google.com/spreadsheets/d/1GMMfhyLK0-q8Xklo3YxIWaZA5vVMuhU1kg41g4xLkXc/edit?hl%3Den_GB%26hl%3Den_GB%23gid%3D1569422585&sa=D&source=docs&ust=1694449661118386&usg=AOvVaw03zagIduGr-OIWcQacdwh.
- Villain, A. (2018) *Étude génomique des interactions diatomées-bactéries*. PhD Thesis. Disponible sur: <http://www.theses.fr/2018AIXM0202/document>.
- Vilmi, A. et al. (2015) 'Freshwater diatoms as environmental indicators: evaluating the effects of eutrophication using species morphology and biological indices', *Environmental Monitoring and Assessment*, 187(5), p. 243. Disponible sur: <https://doi.org/10.1007/s10661-015-4485-7>.
- Vincent, F. and Bowler, C. (2022) 'An Integrated View of Diatom Interactions', in A. Falciatore and T. Mock (eds) *The Molecular Life of Diatoms*. Cham: Springer International Publishing, pp. 59–86. Disponible sur: https://doi.org/10.1007/978-3-030-92499-7_3.

- Vurture, G.W. et al. (2017) 'GenomeScope: fast reference-free genome profiling from short reads', *Bioinformatics*. Edited by B. Berger, 33(14), pp. 2202–2204. Disponible sur: <https://doi.org/10.1093/bioinformatics/btx153>.
- Waksman, S.A. and Butler, M.R. (1937) 'Relation of bacteria to diatoms in sea water', *Journal of the Marine Biological Association of the United Kingdom*, 22(1), pp. 359–373.
- Wang, F. et al. (2018) 'Assessment of 16S rRNA gene primers for studying bacterial community structure and function of aging flue-cured tobaccos', *AMB Express*, 8(1), p. 182. Disponible sur: <https://doi.org/10.1186/s13568-018-0713-1>.
- Wang, Y. et al. (2014) 'Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis', *PLoS ONE*. Edited by N.C.M. Gomes, 9(3), p. e90053. Disponible sur: <https://doi.org/10.1371/journal.pone.0090053>.
- Wang, Yunhao et al. (2021) 'Nanopore sequencing technology, bioinformatics and applications', *Nature Biotechnology*, 39(11), pp. 1348–1365. Disponible sur: <https://doi.org/10.1038/s41587-021-01108-x>.
- Watanabe, T. (1963) 'INFECTIVE HEREDITY OF MULTIPLE DRUG RESISTANCE IN BACTERIA', *Bacteriological Reviews*, 27(1), pp. 87–115. Disponible sur: <https://doi.org/10.1128/br.27.1.87-115.1963>.
- Watkins, R.R. and Bonomo, R.A. (2016) 'Overview: Global and Local Impact of Antibiotic Resistance', *Infectious Disease Clinics of North America*, 30(2), pp. 313–322. Disponible sur: <https://doi.org/10.1016/j.idc.2016.02.001>.
- Watson, J.D. and Crick, F.H. (1953) 'Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid', *Nature*, 171(4356), pp. 737–738.
- Wells, J.N. and Feschotte, C. (2020) 'A Field Guide to Eukaryotic Transposable Elements', *Annual Review of Genetics*, 54(1), pp. 539–561. Disponible sur: <https://doi.org/10.1146/annurev-genet-040620-022145>.
- West, P.T. et al. (2018) 'Genome-reconstruction for eukaryotes from complex natural microbial communities', *Genome Research*, 28(4), pp. 569–580. Disponible sur: <https://doi.org/10.1101/gr.228429.117>.
- Wetzel, C.E. et al. (2012) 'Tursiocola podocnemicola sp. nov., a new epizoic freshwater diatom species from the Rio Negro in the Brazilian Amazon Basin', *Diatom Research*, 27(1), pp. 1–8. Disponible sur: <https://doi.org/10.1080/0269249X.2011.642498>.
- Wheatley, R.M. et al. (2022) 'Gut to lung translocation and antibiotic mediated selection shape the dynamics of *Pseudomonas aeruginosa* in an ICU patient', *Nature Communications*, 13(1), p. 6523. Disponible sur: <https://doi.org/10.1038/s41467-022-34101-2>.
- Wick, R.R. and Holt, K.E. (2019) 'Benchmarking of long-read assemblers for prokaryote whole genome sequencing', *F1000Research*, 8.

- Wijewardene, L. et al. (2022) 'Epiphytic biofilms in freshwater and interactions with macrophytes: Current understanding and future directions', *Aquatic Botany*, 176, p. 103467. Disponible sur: <https://doi.org/10.1016/j.aquabot.2021.103467>.
- Wood, D.E., Lu, J. and Langmead, B. (2019) 'Improved metagenomic analysis with Kraken 2', *Genome Biology*, 20(1), p. 257. Disponible sur: <https://doi.org/10.1186/s13059-019-1891-0>.
- World Health Organization (2007) *The world health report 2007: a safer future: global public health security in the 21st century*. World Health Organization.
- World Health Organization (2014) *Antimicrobial resistance: global report on surveillance*. World Health Organization.
- Wu, J. et al. (2023) 'Antibiotics and antibiotic resistance genes in agricultural soils: A systematic analysis', *Critical Reviews in Environmental Science and Technology*, 53(7), pp. 847–864.
- Yoon, H.S. et al. (2002) 'The Single, Ancient Origin of Chromist Plastids', *Journal of Phycology*, 38(s1), pp. 40–40. Disponible sur: <https://doi.org/10.1046/j.1529-8817.38.s1.8.x>.
- Yorki, S. et al. (2023) 'Comparison of long- and short-read metagenomic assembly for low-abundance species and resistance genes', *Briefings in Bioinformatics*, 24(2), p. bbad050. Disponible sur: <https://doi.org/10.1093/bib/bbad050>.
- Youle, M. et al. (2013) 'Superorganisms and Holobionts: Looking for a term for the functional entity formed by a macrobe and its associated symbiotic microbes and viruses? The term is "holobiont"', *Microbe Magazine*, 8(4), pp. 152–153. Disponible sur: <https://doi.org/10.1128/microbe.8.152.1>.
- Zahajská, P. et al. (2020) 'What is diatomite?', *Quaternary Research*. 2020/04/29 edn, 96, pp. 48–52. Disponible sur: <https://doi.org/10.1017/qua.2020.14>.
- Zebiri, I. et al. (2023) 'The Polysaccharidic Nature of the Skeleton of Marennine as Determined by NMR Spectroscopy', *Marine Drugs*, 21(1), p. 42. Disponible sur: <https://doi.org/10.3390/md21010042>.
- Zerbino, D.R. and Birney, E. (2008) 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs', *Genome research*, 18(5), pp. 821–829.
- Zhang, X. et al. (2022) 'Benchmarking of long-read sequencing, assemblers and polishers for yeast genome', *Briefings in Bioinformatics*, 23(3), p. bbac146. Disponible sur: <https://doi.org/10.1093/bib/bbac146>.

Annexes

1. Tentatives approfondies de décontamination du métagénome

1) Analyses des contigs > 1Mb

Une première identification taxonomique a été réalisée sur les contigs assemblés et possédant une taille supérieure ou égale à 1Mb. Les caractéristiques de ces 24 dernières ont été récupérées et mises en forme dans un tableau (Tableau 1-3) : identifiant, taille, %GC, circularité. La composition de ces contigs apparaît alors très diverse, les %GC variant de 45% à 60%. Également, la moitié d'entre eux étaient circularisés. Ces observations pourraient indiquer que la majorité de ces contigs proviendraient d'un procaryote ou d'un plasmide, le génome circulaire étant souvent associé à ce type de microorganismes, et ce indépendamment du fait que le %GC soit proche de celui attendu pour *H. ostrearia*. Historiquement, les bactéries sont connues pour posséder un %GC variant entre 25% et 75% (Sueoka 1962), voire même plus bas avec un 20%GC pour le génome de *Carsonella* (Mann et Chenn 2010). Cependant, il est nécessaire d'annoter taxonomiquement ces derniers avant de formuler toute hypothèse.

Deux méthodes ont alors été employées pour identifier leur provenance. La première a été de rechercher la présence de gènes codant pour les ARNr 16S. Pour cela, plusieurs amorces classiquement utilisées par la communauté scientifique (Wang et al. 2018) pour une PCR 16S ont été récupérées et alignées par un Blastn (Altschul et al 1990) sur nos contigs. L'option `-word_size 6` a été précisée pour cet alignement comme les séquences d'amorces de PCR ont tendance à être courtes (entre 17 et 25b). Blast commençant par la recherche d'une séquence parfaite correspondant à la longueur donnée par `-word_size`, il a donc été nécessaire de réduire celle-ci à 6b (par défaut à 11b). La recherche de gènes ARNr 18S n'a produit aucun résultat, les amorces utilisées étant celles décrites par Wang et al (2014). Un assignement taxonomique pour les différentes séquences récupérées a alors été proposé en

utilisant la base de données 16S SILVA 132.99 (Quast et al 2012) et le logiciel Qiime2 2021.2.0 (Bolyen et al 2019). Le meilleur résultat de cet assignement et le coefficient de confiance associé à chaque séquence est résumé dans le Tableau 1-3 (assignation taxonomique la plus longue pour un coefficient de confiance > 0,5). Seuls 4 contigs sur les 24 récupérés n'ont ni été détectés par 16S et ni par 18S, tous les autres ont été identifiés comme provenant d'un génome bactérien. Cette méthode a d'ailleurs également été utilisée pour étudier d'autres travaux réalisées sur un second isolat de *H. ostrearia*, dont les résultats sont présentés dans la partie 2.4 de l'article de Gabed et al (2022).

La deuxième méthode a utilisé le logiciel BUSCO (Manni et al 2021) pour étudier la complétude de chacun de ces contigs pour un organisme donné en se basant sur le contenu en gènes orthologues. Si le contig considéré obtient un score BUSCO important (>80%) pour un organisme donné (eucaryotes ou procaryotes), alors on peut considérer que ce dernier est complet, ou quasiment, pour celui-ci et qu'il n'y a pas de doute à avoir sur sa provenance. Cependant, un certain recul concernant la base de données associée aux algues a été pris : en effet, il n'a pas été écarté l'idée que le génome de *H. ostrearia* pouvait être plus grand que les contigs obtenus ici. Le score BUSCO s'en retrouverait donc diminué par la possibilité que le contig ne contiendrait qu'une partie des gènes normalement partagés par toutes les algues. Un à un, les contigs ont donc été comparés à la base de données des *Stramenopiles* (algues) et à la base de données correspondant à l'association taxonomique réalisée précédemment (Tableau 1-3). Les 4 contigs non identifiés par la méthode 16S ont, quant à eux, été comparés à la base de données générale des bactéries. Le score BUSCO associé à une base de données procaryotes a été retrouvé supérieur à 80% pour 14 contigs sur 24, seuls 6 contigs ont obtenu un score BUSCO inférieur à 50%. L'identification taxonomique obtenue précédemment a donc été confirmée pour ces 14 premiers, et ils ont été classifiés comme génomes bactériens entiers ou quasiment, et ce quelque soit le score obtenu face à la base de données des algues. Pour les 6 autres, ce score a été pris en considération pour conclure sur leur identification taxonomique. Seul le contig #824 avait dépassé les 5%, avec 6 gènes complets orthologues des algues (6%). Malheureusement la méconnaissance des

Magnetospiraceae nous a obligé à utiliser la base de données des *Alphaproteobacteria*, il est possible que cela soit la raison d'un score aussi faible pour un organisme procaryote. La même observation a été faite pour le contig #228 et les *Cryomorphaceae* vis à vis de la base de données des *Flavobacteriales*. L'appartenance taxonomique étant souvent revue au fil des années (Bowman 2020, Koziava et al 2023), il est possible que la base de données sélectionnée ne soit pas celle à utiliser. De plus, au vu de la présence d'un gène 16S chez #824 et #228, il était peu probable que ces dernières proviennent de *H. ostrearia*. Ce n'était pas le cas des contigs #1265, #767, #25 et #1806 pour lesquels aucun gène 16S n'a été retrouvé. Cependant, le test du score BUSCO face aux *Stramenopiles* n'a pas pu être réalisé pour #1265, le logiciel refusant tout calcul pour une raison inconnue. Seul #767 contenait 4 gènes complets et orthologues des algues et pu être soupçonné d'appartenir à *H. ostrearia*. Mais aucune réelle conclusion n'a donc pu être proposée pour ces 4 contigs. Les contigs restants ont été classifiés comme bactériens, avec de forts soupçons, mais sont restés considérés comme des contigs incomplets et non des génomes.

Face à ces résultats, il est clair que les séquences de *H. ostrearia* dans le métagénome ont été difficiles à assembler et qu'elles seraient assez courtes. De plus, de nombreuses bactéries seraient également présentes dans les données de séquençage rendant la tâche plus difficile encore. Enfin, il a été remarqué que ces dernières ne seraient pas seulement présentes dans les données de séquençage associées à un %GC de 60%, comme l'ont prouvé celles identifiées précédemment (Tableau 1-3). Des analyses plus poussées sont donc nécessaires pour identifier les séquences appartenant à la microalgue, notamment dans les contigs de petite taille.

Paire d'amorces	Séquence (5' à 3')	Région 16S (V)	Référence
8F	TGGAGAGTTTGATCCTGGCTCAG	V1V2V3	Sun et al (2014)
533R	TACCGCGGCTGCTGGCAC		
336F	GTACTCCTACGGGAGGCAGCA	V3V4	Munyaka et al (2015)
806R	GTGGACTACHVGGGTWTCTAAT		
515F	GTGCCAGCMGCCGCGGTAA	V4V5	Tuan et al (2014)
909R	CCCCGYCAATTCMTTTRAGT		
799F	AACMGGATTAGATACCKG	V5V6V7	Beckers et al (2016)
1193R	ACGTCATCCCCACCTTCC		

Tableau 1-2 : Liste des amorces utilisées pour la recherche in silico des gènes ARNr 16S dans le métagénome de *Haslea ostrearia*, Tableau traduit de Wang et al (2018)

Les amorces sont indiquées comme étant directes (F) ou inverses (R). Les régions 16S (V) sont les régions hypervariables de l'opéron ARNr 16S ciblées par les paires d'amorces.

(A)

Contig	Longueur (bases)	%GC	Circulaire ?	Taxon identifié par Qiime2	Confiance
contig_357	7116910	45,38%	non	<i>D_5__Aestuariatbacter;D_6__uncultured bacterium</i>	0.8
contig_300	5333463	41,26%	non	<i>D_5__Fluviicola</i>	0.9
contig_182	4866340	39,80%	non	<i>D_4__Cyclobacteriaceae</i>	1.0
contig_1381	4637382	52,33%	oui	<i>D_2__Gammaproteobacteria;D_3__OM182 clade;D_4__uncultured bacterium;D_5__uncultured bacterium;D_6__uncultured bacterium</i>	0.6
contig_93	4351421	62,26%	oui	<i>D_5__Phycisphaera</i>	0.9
contig_87	4231418	57,70%	non	<i>D_5__Marivita</i>	0.9
contig_824	4054853	52,66%	non	<i>D_4__Magnetospiraceae;D_5__uncultured</i>	0.9
contig_4	3983877	60,14%	oui	<i>;D_5__Salinhabitans</i>	0.7
contig_660	3655487	57,91%	oui	<i>D_5__Loktanella</i>	0.8
contig_344	3622812	38,66%	non	<i>D_5__Roseivirga</i>	0.8
contig_650	3556514	58,45%	oui	<i>D_5__Alcanivorax;D_6__uncultured bacterium</i>	0.7
contig_2057	3524939	55,53%	oui	<i>D_4__Hyphomonadaceae;D_5__uncultured</i>	1.0
contig_90	3463066	65,19%	oui	<i>D_5__Cyanobium PCC-6307</i>	1.0
contig_1735	3449784	57,26%	non	<i>D_3__Rhizobiales</i>	0.7
contig_473	3444470	58,84%	oui	<i>D_2__Alphaproteobacteria</i>	1.0
contig_91	3313448	61,50%	oui	<i>D_5__Roseovarius;Ambiguous_taxa</i>	0.6
contig_228	3148984	36,93%	oui	<i>D_4__Cryomorphaceae;D_5__uncultured;D_6__uncultured bacterium</i>	0.8
contig_101	2970055	61,91%	oui	<i>D_5__Oceanicaulis</i>	0.7
contig_1095	2963715	61,10%	oui	<i>D_4__Rhodobacteraceae</i>	1.0
contig_713	2746463	56,60%	non	<i>D_3__Cellvibrionales</i>	1.0
contig_1265	1339044	43,32%	non	N/A	N/A
contig_767	1157211	43,76%	non	N/A	N/A
contig_25	1087446	60,34%	non	N/A	N/A
contig_1806	1082056	59,79%	non	N/A	N/A

(B)

Contig	BUSCO vs Base de données Procaryotes			BUSCO vs Base de données Eucaryotes (<i>Stramenopiles</i>)	
	Base de données utilisée	Nombre de gènes retrouvés (/ Nombre de gènes recherchés)	Score (%)	Nombre de gènes retrouvés(/ Nombre de gènes recherchés)	Score (%)
contig_357	<i>Alteromonadales</i>	762 complets (/820)	92,9%	15 complets (/100)	15,0%
contig_300	<i>Flavobacteriales</i>	662 complets (/733)	90,3%	7 complets (/100)	7,0%
contig_182	<i>Bacteroidetes</i>	365 complets (/402)	90,8%	2 complets (/100)	2,0%
contig_1381	<i>Bacteria</i>	119 complets (/124)	96,0%	13 complets (/100)	13,0%
contig_93	<i>Bacteria</i>	103 complets (/124)	83,1%	10 complets (/100)	10,0%
contig_87	<i>Rhodobacterales</i>	776 complets (/833)	93,1%	12 complets (/100)	12,0%
contig_824	<i>Alphaproteobacteria</i>	196 complets (/432)	45,4%	6 complets (/100)	6,0%
contig_4	<i>Rhodobacterales</i>	755 complets (/833)	90,6%	9 complets (/100)	9,0%
contig_660	<i>Rhodobacterales</i>	809 complets (/833)	97,1%	10 complets (/100)	10,0%
contig_344	<i>Bacteroidetes</i>	259 complets (/402)	64,4%	7 complets (/100)	7,0%
contig_650	<i>Oceanospirillales</i>	588 complets (/619)	95,0%	13 complets (/100)	13,0%
contig_2057	<i>Alphaproteobacteria</i>	337 complets (/432)	78,0%	10 complets (/100)	10,0%
contig_90	<i>Synechococcales</i>	667 complets (/788)	84,7%	7 complets (/100)	7,0%
contig_1735	<i>Rhizobiales</i>	600 complets (/639)	93,9%	12 complets (/100)	12,0%
contig_473	<i>Rhizobiales</i>	615 complets (/639)	96,3%	14 complets (/100)	14,0%
contig_91	<i>Rhodobacterales</i>	622 complets (/833)	74,6%	10 complets (/100)	10,0%
contig_228	<i>Flavobacteriales</i>	216 complets (/733)	29,4%	1 complets (/100)	1,0%
contig_101	<i>Alphaproteobacteria</i>	409 complets (/432)	94,6%	12 complets (/100)	12,0%
contig_1095	<i>Alphaproteobacteria</i>	381 complets (/432)	88,2%	N/A	N/A
contig_713	<i>Cellvibrionales</i>	603 complets (/913)	66,0%	8 complets (/100)	8,0%
contig_1265	<i>Bacteria</i>	35 complets (/124)	28,2%	N/A	N/A
contig_767	<i>Bacteria</i>	22 complets (/124)	17,7%	4 complets (/100)	4,0%
contig_25	<i>Bacteria</i>	6 complets (/124)	4,8%	0 complets (/100)	0,0%
contig_1806	<i>Bacteria</i>	18 complets (/124)	14,5%	1 complets (/100)	1,0%

Tableau 1-3 : Informations sur les résultats obtenus pour l'analyse taxonomique des contigs dépassant la taille de 1 Mb et appartenant à l'assemblage métagénomique des cultures de *Haslea ostrearia*

Les résultats présentés ici ont été obtenus de l'assignation taxonomique des régions 16S retrouvées *in silico* par Qiime2 2021.2.0 (Bolyen et al 2019) contre la base de données SILVA 132.99 (Quast et al 2012) (A) et lors de l'analyse par BUSCO (Manni et al 2021) des contigs (B).

La base de données procaryote à utiliser pour l'analyse (B) a été sélectionnée manuellement et correspond à la classe taxonomique identifiée précédemment lors de l'analyse (A), afin de pouvoir comparer les résultats entre eux.

2) Recherche des contigs appartenant à *H. ostrearia* dans le métagénome

La première tentative pour identifier les séquences d'intérêt dans notre métagénome a été de proposer une annotation taxonomique pour l'ensemble des contigs. Pour cela, le logiciel Kraken2 2.0.9beta (Wood et al 2019) a été choisi. Il s'agit d'un logiciel spécialement conçu pour classer rapidement et précisément des séquences ADN à l'aide de k-mers. Ce dernier utilise sur une base de données contenant, entre autres, quelques informations sur l'embranchement des *Bacillariophyta* (diatomées) mais, surtout, sur *H. ostrearia*. Ainsi, 1040 contigs ayant été labellisés comme procaryotes par Kraken2 ont été annotés comme contaminants dans nos données, les 21 labellisés comme appartenant à un membre des *Bacillariophyta* ont été marqués comme potentiellement provenant de *H. ostrearia*, et enfin les 9 marqués comme *H. ostrearia* ont été annotés comme tels. Les 1553 autres contigs ont été labellisés comme "unclassified" par Kraken2 et n'ont pas fait l'objet d'une modification de leur statut dans nos données. Les raisons derrière cet important nombre de séquences non identifiées pourraient être un biais dans les données utilisées par Kraken2, puisqu'en effet elles sont connues pour être composées en majorité de bactéries retrouvées dans les intestins d'êtres humains, ou la taille des contigs à analyser, beaucoup de petits contigs dans le métagénome, ou encore la non spécificité des k-mers qui leur sont associés, séquences communes à trop d'espèces différentes.

Afin d'ajouter du poids à cette annotation proposée par Kraken2, les contigs associés aux *Bacillariophyta* ou à notre microalgue ont été comparés à la base de données nt du NCBI par un Blastn (Altschul et al 1990). Si la séquence s'alignait contre celle d'une diatomée avec un %ID et un %couverture supérieures ou égales à 90, alors le contig a été marqué comme provenant de la microalgue. 12 contigs ont ainsi été identifiés comme appartenant à *H. ostrearia* (Tableau 1-4), les autres ont été laissés dans la liste des "contigs suspectés".

La deuxième tentative a été de pallier à l'absence de génome de référence de *H. ostrearia* en utilisant directement les données disponibles pour les diatomées les mieux

connues. Pour cela, nous avons récupéré de la base de données du NCBI les séquences codantes (CDS) de *T. pseudonana* et ceux de *P. tricornutum*, puis les avons alignés via un Blastn sur le métagénome. Seules les séquences alignées avec un pourcentage d'identité (%ID) supérieur à 80 ont été considérées pour la suite. Respectivement, 130 et 145 contigs ont été reconnus comme ayant des similarités avec les CDS des diatomées connues. Toutefois, comme ces CDS ne sont pas forcément spécifiques aux diatomées, d'autres filtres ont également été mis en place pour renforcer nos soupçons sur leur appartenance à *H. ostrearia*. Ainsi, nous avons également vérifié si ces contigs étaient proches des CDS appartenant aux deux diatomées ou seulement à une d'entre elles. Ce filtre a permis de réduire le nombre de contigs hypothétiques à 64.

Enfin, devant le manque d'informations sur *H. ostrearia* et sa contamination évidente, la plateforme de séquençage nous a proposé d'opter pour un séquençage Illumina des ARNm possédant une queue polyA comme dernier filtre pour identifier les séquences de la microalgue (Chapitre 1-1 - Tableau 1). L'idée était d'utiliser le fait qu'il existe peu de bactéries capables de produire ces ARNm, il y avait donc de fortes probabilités que ceux récupérés appartiendraient à *H. ostrearia*. En les alignant sur le métagénome, en prenant en considération la présence d'introns lors de l'alignement sur l'ADN, alors nous pourrions identifier précisément ses séquences. Ainsi, les ARNm polyA séquencés en Illumina ont été filtrés comme précédemment (Chapitre 1-2.1) et assemblés de manière *de novo* en utilisant Trinity 2.11.0 (Haas et al 2013). Grâce à STAR 2.7.5a (Dobin et al 2013), les 109 932 séquences obtenues ont été alignées sur le génome en prenant en considération la potentielle présence d'introns. Après observation des résultats obtenus, il a été mis en évidence que beaucoup de ces ARNm polyA s'alignaient parfaitement (%ID = 100 et un %couverture = 100) sur des contigs avec une taille supérieure à 1Mb. Pourtant il s'agissait de contigs précédemment détectés comme étant bactériens ! Des recherches plus approfondies ont alors révélé que la proposition de la plateforme de séquençage était erronée : les bactéries sont en possession d'ARNm avec une queue polyA. La connaissance sur leur existence est cependant encore peu répandue, puisqu'il a longtemps été considéré que seuls les eucaryotes en

possédaient (Hajnsdorf et Kaberdin 2018). Afin de vérifier l'appartenance de ces ARNm polyA et leur complétude, un score BUSCO (Manni et al 2021) a été calculé en confrontant ce transcriptome à la base de données des *Stramenopiles* (algues) et celle des bactéries. Le transcriptome comprenait quasiment l'entièreté des gènes partagés par les *Stramenopiles* (97% des 100 gènes recherchés) mais également 33.95% de ceux bactériens (~34% de 124 gènes). Bien que le risque d'avoir également récupéré des ARN polyA de bactéries ne peut être complètement écarté, il semble néanmoins que les données utilisées ici comprenaient de nombreux ARN polyA provenant d'algues (au moins 100) et peu de bactéries (<42). Il a donc été choisi de ne garder que les contigs pour lesquels l'alignement contre un ARN polyA comprenait un gap d'au minimum 60 bases pour la séquence ARN, signe de la potentielle présence d'un intron, et donc un nouveau soupçon d'une appartenance à *H. ostrearia* puisqu'il s'agit d'une caractéristique seulement partagée par les eucaryotes. Malheureusement, la présence de gaps pouvait également être le signe d'un morceau d'ADN supplémentaire n'ayant pas de lien avec un intron, soit une séquence ADN ressemblant à celle de l'ARN polyA sur deux régions éloignées. C'est pourquoi il a semblé nécessaire de n'utiliser cette méthode que sur les 64 contigs précédemment identifiés. Ainsi, 31 contigs ont été listés comme ayant de forts soupçons d'appartenir à *H. ostrearia* (Tableau 1-4).

Bien que peu de contigs aient été identifiés grâce à ces deux méthodes, il était clair qu'une tendance se dessinait (Tableau 1-4). Premièrement, le %GC des séquences obtenues semblaient correspondre au 45%GC, soit celui soupçonné en premier lieu lors de la visualisation des données brutes. Seuls les génomes mitochondrial et chloroplastique dérogeaient à cette observation avec des %GC plus proche des 30%. Deuxièmement, la profondeur de séquençage associée à chaque séquence semblait très faible pour la majorité de celles-ci (~10X), sauf pour quelques exceptions comme les contigs identifiés comme la mitochondrie et le chloroplaste de *H. ostrearia*. En comparaison, celle observée pour la plupart des contigs bactériens dépassait très largement les 100X. Dernièrement, la taille des contigs se situaient, sauf exceptions, en dessous des 50kb.

Comme attendu, la faible profondeur et la contamination très présente semblent avoir rendu la reconstruction du génome nucléaire de *H. ostrearia* difficile. Cependant, le génome de la mitochondrie et celui du chloroplaste semblent avoir été bien assemblés et identifiés.

Enfin, d'autres méthodes ont été tentées et sont résumées ci-dessous. Les résultats n'ont pas été ceux escomptés puisqu'ils ont été soit impossible à interpréter, soit erronés.

Il a été essayé de retrouver les contigs appartenant à un eucaryote en alignant le métagénome à la base de données RepBase datant de 26/10/2018 (Jurka et al 2005) via RepeatMasker 4.1.2.p1 (<http://www.repeatmasker.org>), plus précisément aux retrotransposons contenus dans cette base de données (dont ceux des plantes), puisque, en effet, seuls les eucaryotes possèdent cette catégorie d'éléments transposables (Eickbush 2001). Cependant, des alignements ont été possible même sur des contigs déjà associés à un taxon bactérien, et il a donc été impossible d'exploiter les résultats pour identifier des contigs eucaryotes. La même observation a été faite avec la base de données redondantes V2 de retrotransposons de plantes de Inpactor (Orozco-Arias et al 2018) et le logiciel d'alignement Blastn.

Il a également été tenté de ne récupérer que les contigs associés à une profondeur avoisinant celles obtenues pour les contigs identifiés comme étant de *H. ostrearia*. Malheureusement le peu d'information génétique contenu dans ces derniers n'a pas permis d'identifier s'il s'agissait de bactéries rares ou de microalgue (contig de petites tailles).

Le logiciel EukRep a également été essayé sur les données du métagénome, puisqu'il se présente comme étant conçu pour classifier des séquences eucaryotes et procaryotes à partir d'un jeu de données métagénomiques (West et al 2018). Malheureusement les données devaient être trop hétéroclites ou les génomes bactériens trop proches du génome de *H. ostrearia*, car le logiciel n'a pas su reconnaître de contigs eucaryotes dans notre métagénome.

Enfin, l'hypothèse que les bactéries avec un %GC autour de celui identifié comme étant celui de la microalgue (~45%GC) devaient être en plus petites quantités a également été émise. Il suffisait donc de n'assembler que les lectures ayant un %GC entre 40% et 50% inclus

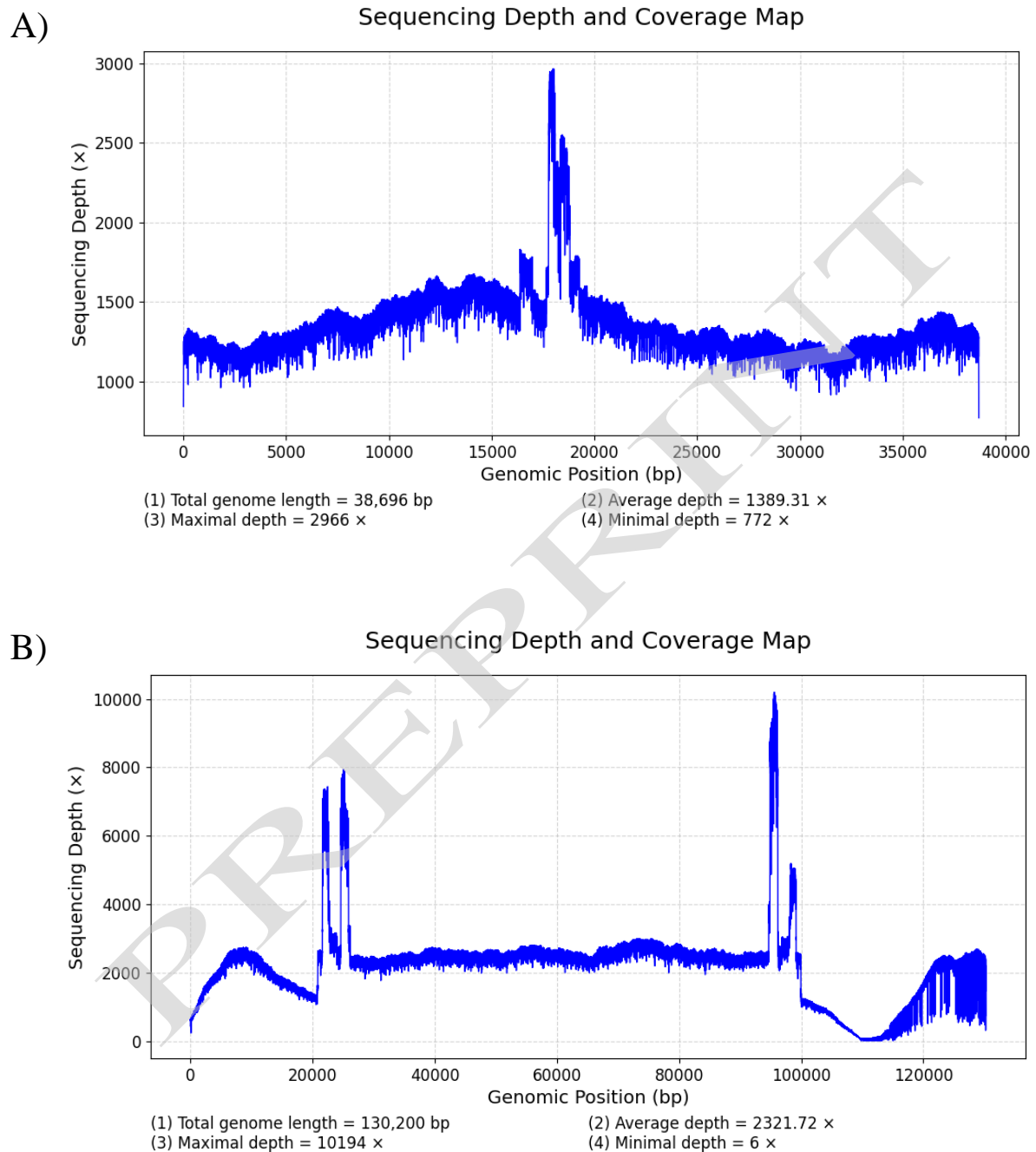
puis d'identifier, dans cet assemblage en partie filtré des bactéries, les séquences provenant de *H. ostrearia*. Le nombre de bactéries est en réalité assez important et le séquençage ne contenait pas de séquences de *H. ostrearia* plus intéressantes que précédemment (plus longues ou plus facilement identifiables par les méthodes précédentes). Cette méthode a été testée en automatique et en bouclant sur l'étape d'assemblage des lectures à ~45%GC et de décontamination des données : les lectures ne s'alignant pas sur les contigs détectés comme bactérien dans un assemblage T ont été recherchées pour reformer un assemblage T+1, et ceci jusqu'à atteindre T=5.

Méthode	Contig	Longueur (bases)	%GC	Profondeur de séquençage (Nanopore)
(A) Kraken2 avec confiance à 0,0 + Blastn avec %id ≥ 90 et %couverture ≥ 90	contig_126	41536	45,26%	7,2
	contig_1314	18975	43,66%	5,1
	contig_1643	28089	45,98%	6,9
	contig_1883 (Mitochondrie)	38695	28,67%	215,8
	contig_208	62630	45,77%	306,1
	contig_2180	11920	46,59%	48,0
	contig_2513 (Chloroplaste)	170851	30,79%	187,7
	contig_2537	30769	44,3%	7,2
	contig_2687	34558	43,79%	5,0
	contig_2794	32275	44,66%	7,2
	contig_359	52930	46,15%	8,8
	(B) Analyse transcrits ARNpolyA + CDS diatomées + %GC + Profondeur Nanopore	contig_1056	29862	45,32%
contig_1067		33101	43,41%	49,18
contig_1156		44861	43,03%	22,93
contig_1206		108701	44,14%	14,46
contig_1240		32708	43,23%	6,06
contig_1263		40044	42,47%	7,72
contig_1306		40315	44,59%	7,58
contig_1314		18975	43,66%	5,10
contig_1336		41758	43,5%	13,23
contig_1606		47462	42,8%	5,97
contig_1947		58193	44,24%	7,19
contig_1998		43111	44,4%	11,64
contig_2044		84168	53,81%	6,64
contig_2113		17777	41,36%	6,98
contig_2206		23618	43,2%	11,37
contig_2353		31556	43,41%	5,96
contig_2385		22487	44,07%	5,79
contig_2508		25039	44,23%	5,81
contig_262		24811	43,77%	8,72
contig_2760		17638	44,2%	7,73
contig_2771		25979	44,11%	4,62
contig_295		45956	44,4%	4,80
contig_458		40933	45,33%	6,67
contig_471		33719	43,15%	9,73
contig_511		35274	44,19%	7,48
contig_759		20815	42,55%	6,45
contig_808		18277	41,87%	13,63
contig_829		44903	42,8%	12,09
contig_922		22434	41,88%	4,70
contig_94		35176	44,21%	6,91
contig_970	20117	43,84%	7,23	

Tableau 1-4 : Informations sur les séquences identifiées comme susceptibles d'appartenir au génome de *Haslea ostrearia*

Différents filtres ont été utilisés pour obtenir ces résultats. (A) Une recherche de séquences homologues par les logiciels Kraken2 2.0.9beta (Wood et al 2019) et Blastn (Altschul et al 1990). Les contigs du métagénome identifiés comme semblables à des séquences de *Bacillariophyta* (diatomées) avec, pour Blast, un pourcentage d'identité de séquences (%id) et un pourcentage de couverture (%couverture) supérieurs ou égal à 90%, et une confiance à 0,0 pour Kraken2, ont été marqués comme appartenant à *Haslea ostrearia* et leurs informations ajoutées au tableau présenté ici. (B) De même pour les contigs validant plusieurs des analyses suivantes : au moins un alignement de transcrits d'ARN poly-A avec présence d'introns, au moins un CDS de diatomées retrouvé, un pourcentage de Guanine-Cytosine (%GC) autour des 45% et/ou une profondeur de séquençage par Nanopore (Prof(ONT)) inférieure à 15X.

2. Supplementary data of “Complete mitochondrial genome and draft chloroplastic genome of *Haslea ostrearia* (Simonsen 1974)”



Supplementary Figure 1: Pacbio read coverage depth map of the mitochondrial genome (A) and the chloroplastic genome (B) of *H. ostrearia*

1 **3. Supplementary data of “To improve the axenic culture protocol,**
2 **the bacteria present in *Haslea ostrearia* cultures need to be**
3 **studied”**

4 **Detailed material and methods**

5 The isolate was registered under the reference NCC 532 in the Nantes Culture Collection (Prof. Vona
6 Meleder, vona.meleder@univ-nantes.fr; Nantes, France). The four cultures of this isolate were grown in
7 an enriched artificial sea water (Instant Ocean, Aquarium systems Ø ; Harrison et al. 1980 [1] modified
8 by De Brouwer et al. 2002 [2]). They were maintained at a constant temperature of 14°C and under a
9 light/dark cycle of 14h/10h under 300m photons/m²/s. Per mL, the antibiotic used contained 10,000 units
10 of penicillin, 10 mg of streptomycin and 25µg of amphotericin B. On the 21th day of growth, the biomass
11 was collected by filtration (Whatman™ Binder-Free Glass Microfiber Filters, Grade GF/C, pore size of
12 1.2 µm). The whole DNA was extracted with the method of Puppo et al. 2017 [3]. The 16S regions were
13 selected with the primers 8F TGGAGAGTTTGATCCTGGCTCAG - 533R
14 TACCGCGGCTGCTGGCAC (V1V3) [4], and 336F GTACTCCTACGGGAGGCAGCA - 806R
15 GTGGACTACHVGGGTWTCTAAT (V3V4) [5].

16 The quality of the obtained reads were viewed by fastqc 0.11.8 and multiqc 1.7 [6, 7]. To remove the
17 primers still present in the data, the first 9 bases of each sequence were deleted by cutadapt 4.0. Pairs with
18 a read that didn't meet a Q quality criterion of more than 20 were also removed during this process:
19 cutadapt -u 9 -U 9 -m 250 -q 20 --pair-filter=both --trim-n [8]. In order to recover the whole amplicons,
20 the filtered reads were then assembled using the Matam software (Apr2020) with its default settings [9].
21 The sequences used thereafter came from the generated scaffolds file, except for the Mock community
22 samples where the non redundant contigs file was picked. Qiime2 2021.2.0 and the SILVA 16S database
23 138 Ref NR 99 were then utilized for the taxonomic identification [10, 11]. The only setting changed in
24 the Qiime2 protocole was the selectionned feature classifier: classify-consensus-blast. To calculate the
25 mean depth associated with each sequence, the first step was to align the filtered reads on the
26 contigs/scaffolds with bwa-mem2 in default mode [12]. Secondly, the depth at each position of each

27 sequence was then acquired with samtools depth 1.10 in default mode [13]. In the end, the mean was
28 calculated and added to the Qiime2 results with a homemade script using python 3.8.10 and panda 1.4.1
29 [14, 15]. To compare the bacterial population between the samples, a χ^2 test was performed using the
30 function `chisq.test()` from the R package `stats` version 3.6.2 [16].

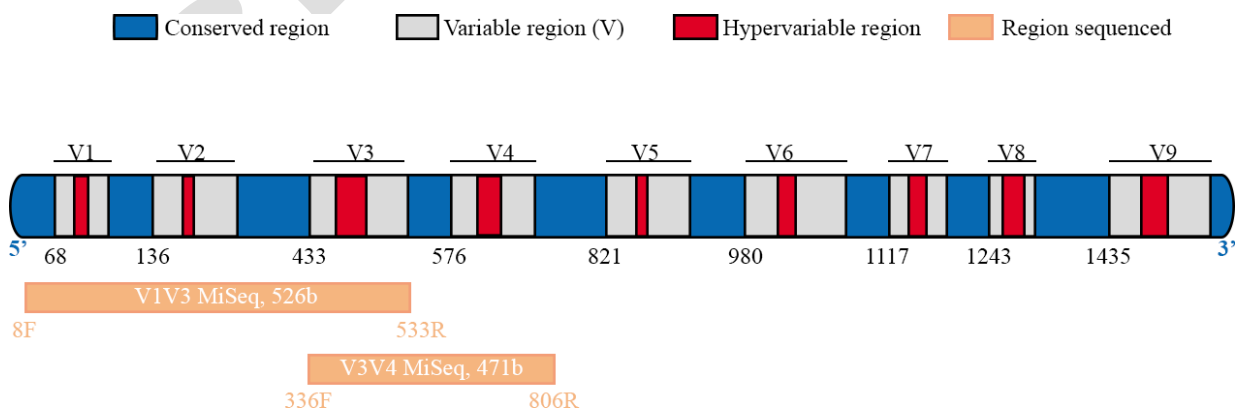
31 All the figures were made under R version 4.2.1 (2022-06-23) using different packages [16]. For the
32 stacked barplot and the grouped barplot `ggplot2` 3.3.5 was used, `dplyr` 1.0.8 for the bubble chart [17, 18].
33 `Reshape2` 1.4.4, `stringR` 1.4.1 and `RColorBrewer` 1.1-3 were also downloaded to reshape the data frame
34 and colorize the plots [19, 20, 21].

35

36 Data summary

37 16S regions V1V3 and V3V4 (Additional fig. 1) were sequenced through Illumina MiSeq platforms
38 (TrueSeqv3, 250pe) for *H. ostrearia* for 2 x 2 types of sample, in laboratory conditions: 2 without (N)
39 and 2 with an antibiotic treatment (T). For each sequencing a Mock community sample was added (Mock,
40 Additional table 1). The final dataset was composed of at least 200,000 150 bases, paired-end (pe)
41 sequences, for all of the samples. After these sequences were assembled, contigs were obtained for the
42 Mock samples and scaffolds for the others.

43



44

45 **Additional figure 1: Illustration of the 16S rRNA regions used for the study on *H. ostrearia* bacteria.**

46 Conserved regions are represented in blue, variable regions in gray and hypervariable regions in red. Primers were
 47 designed in a conserved region to get a PCR product spanning two or more hypervariable regions (V1,V2,V3 and
 48 V3,V4). Double-end (Illumina) sequencing of PCR product provided data for profiling of bacteria of multiple
 49 samples.

50

	V1V3					V3V4				
	N1	N2	T1	T2	Mock	N1	N2	T1	T2	Mock
Sequenced paired of sequences (nb)	215,1 24	353,1 68	346,1 80	361,6 56	201,8 28	209,2 48	203,8 48	300,8 80	295,5 20	319,4 00
Scaffolds or *Contigs assembled (nb)	89	185	171	160	2,413 *	248	257	304	236	22,18 0*
Sequences mapped on the contigs/scaff olds (%)	92.03	94.55	93.35	91.69	92.32	94.57	94.32	94.91	91.32	95.16

51 **Additional table 1: Table containing summarizing information on the data sequenced and used (in**
 52 **black) in the study on *H. ostrearia* bacteria.**

53

54

55 **Samples negative control V3V4**

56 Negative controls were added to the samples in the DNA amplification step. They did not contain any *H.*
57 *ostrearia* DNA but underwent the same manipulations as the other samples. They were used to check for
58 possible human contamination during handling, and to verify the real presence of the bacteria in the *H.*
59 *ostrearia* cultures.

60 The phyla found in the T-neg-V3V4 sample were *Firmicutes* (*Bacilli* 4.53% and *Clostridia* 7.48% of the
61 sample), *Bacteroidetes* (*Bacteroidia* 34.23%), *Actinobacteria* (*Rubrobacteria* 0.7%), *Chloroflexi*
62 (*Dehalococcoidia* 10.8%), *Cyanobacteria* (*Oxyphotobacteria* 3.87%), *Proteobacteria*
63 (*Alphaproteobacteria* 19.87% and *Gammaproteobacteria* 17.76%). An *archaea* phylum was also found:
64 *Thaumarchaeota* (*Nitrososphaeria* 0.75% of the sample). These phyla comprised 11 classes, of which 3
65 of them were identified in T-neg-V3V4 and in *H. ostrearia* V3V4 samples: *Bacilli*, *Oxyphotobacteria*
66 and *Bacteroidia*. But, looking in detail, only the *Chloroplast* order (3.87% of the sample,
67 *Oxyphotobacteria*) and the *Lactobacillus* family (3.87%, *Bacilli*) were common to all the samples.
68 However, they did not have a high abundance in N and T samples (<1% of the sample). As their
69 abundance was negligible in the data examined, they were not considered to contaminate the data and
70 interfere with the analyses. Indeed, the main bacteria considered in this study are the predominant ones
71 and, given the observations made here, do not originate from possible external contamination.

72

73

74

75

Taxonomy V3V4	N1		N2		T1		T2	
	Mean Death	%	Mean Death	%	Mean Death	%	Mean Death	%
<i>D_0_Bacteria</i>	50 568,71	100,00	55 865,89	100,00	85 116,19	100,00	81 871,03	100,00
<i>D_0_Bacteria;D_1_Actinobacteria</i>	0,00	0,00	0,00	0,00	249,89	0,29	183,95	0,22
<i>D_0_Bacteria;D_1_Actinobacteria;D_2_Acidimicrobiia</i>	0,00	0,00	0,00	0,00	105,65	0,12	0,00	0,00
<i>D_0_Bacteria;D_1_Actinobacteria;D_2_Acidimicrobiia;D_3_Microtrichales</i>	0,00	0,00	0,00	0,00	105,65	0,12	0,00	0,00
<i>D_0_Bacteria;D_1_Actinobacteria;D_2_Actinobacteria</i>	0,00	0,00	0,00	0,00	144,24	0,17	183,95	0,22
<i>D_0_Bacteria;D_1_Actinobacteria;D_2_Actinobacteria;D_3_Micrococcales</i>	0,00	0,00	0,00	0,00	144,24	0,17	183,95	0,22
<i>D_0_Bacteria;D_1_Bacteroidetes</i>	8 420,93	16,65	8 800,07	15,75	10 291,97	12,09	9 203,40	11,24
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Bacteroidia</i>	8 420,93	16,65	8 800,07	15,75	10 201,51	11,99	9 148,61	11,17
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Bacteroidia;D_3_Bacteroidales</i>	3 672,62	7,26	2 612,93	4,68	2 261,53	2,66	1 465,64	1,79
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Bacteroidia;D_3_Cytophagales</i>	1 578,82	3,12	467,46	0,84	558,56	0,66	503,96	0,62
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Bacteroidia;D_3_Flavobacteriales</i>	3 012,53	5,96	5 719,68	10,24	7 278,57	8,55	7 100,90	8,67
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Bacteroidia;D_3_Sphingobacteriales</i>	156,96	0,31	0,00	0,00	102,86	0,12	78,10	0,10
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Rhodothermia</i>	0,00	0,00	0,00	0,00	90,47	0,11	54,79	0,07
<i>D_0_Bacteria;D_1_Bacteroidetes;D_2_Rhodothermia;D_3_Balneolales</i>	0,00	0,00	0,00	0,00	90,47	0,11	54,79	0,07
<i>D_0_Bacteria;D_1_Cyanobacteria</i>	5 411,24	10,70	776,84	1,39	776,53	0,91	1 257,16	1,54
<i>D_0_Bacteria;D_1_Cyanobacteria;D_2_Melainabacteria</i>	0,00	0,00	33,78	0,06	33,53	0,04	0,00	0,00
<i>D_0_Bacteria;D_1_Cyanobacteria;D_2_Melainabacteria;D_3_Gastranaerophilales</i>	0,00	0,00	33,78	0,06	33,53	0,04	0,00	0,00
<i>D_0_Bacteria;D_1_Cyanobacteria;D_2_Oxyphotobacteria</i>	5 411,24	10,70	743,06	1,33	743,00	0,87	1 257,16	1,54
<i>D_0_Bacteria;D_1_Cyanobacteria;D_2_Oxyphotobacteria;D_3_Chloroplast</i>	5 411,24	10,70	365,08	0,65	743,00	0,87	1 257,16	1,54
<i>D_0_Bacteria;D_1_Cyanobacteria;D_2_Oxyphotobacteria;D_3_Synechococcales</i>	0,00	0,00	377,99	0,68	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Firmicutes</i>	7 957,27	15,74	6 668,29	11,94	5 236,74	6,15	3 552,53	4,34
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Bacilli</i>	414,29	0,82	335,51	0,60	302,13	0,35	201,63	0,25
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Bacilli;D_3_Lactobacillales</i>	414,29	0,82	335,51	0,60	302,13	0,35	201,63	0,25
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Clostridia</i>	7 163,27	14,17	6 037,35	10,81	4 711,66	5,54	3 181,98	3,89
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Clostridia;D_3_Clostridiales</i>	7 163,27	14,17	6 037,35	10,81	4 711,66	5,54	3 181,98	3,89
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Erysipelotrichia</i>	379,71	0,75	295,42	0,53	222,95	0,26	168,91	0,21
<i>D_0_Bacteria;D_1_Firmicutes;D_2_Erysipelotrichia;D_3_Erysipelotrichales</i>	379,71	0,75	295,42	0,53	222,95	0,26	168,91	0,21
<i>D_0_Bacteria;D_1_Planctomycetes</i>	689,36	1,36	897,04	1,61	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Planctomycetes;D_2_Phycisphaerae</i>	689,36	1,36	897,04	1,61	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Planctomycetes;D_2_Phycisphaerae;D_3_Phycisphaerales</i>	689,36	1,36	897,04	1,61	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Proteobacteria</i>	27 942,59	55,26	38 656,44	69,20	68 320,61	80,27	67 455,24	82,39
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria</i>	19 489,47	38,54	31 454,17	56,30	53 006,34	62,28	52 116,01	63,66

<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Caulobacteriales</i>	5 982,45	11,83	7 265,73	13,01	5 484,02	6,44	4 436,27	5,42
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Parvibaculales</i>	392,28	0,78	0,00	0,00	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Puniceispirillales</i>	102,28	0,20	0,00	0,00	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Rhizobiales</i>	8 536,20	16,88	5 698,39	10,20	11 312,92	13,29	11 291,52	13,79
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Rhodobacteriales</i>	4 406,31	8,71	15 851,40	28,37	33 039,41	38,82	33 018,11	40,33
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Rhodospirillales</i>	0,00	0,00	139,66	0,25	293,68	0,35	240,45	0,29
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Sneathiellales</i>	0,00	0,00	0,00	0,00	0,00	0,00	159,74	0,20
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_Sphingomonadales</i>	0,00	0,00	27,94	0,05	1 007,87	1,18	1 151,49	1,41
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Alphaproteobacteria;D_3_uncultured</i>	0,00	0,00	0,00	0,00	63,34	0,07	60,24	0,07
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Deltaproteobacteria</i>	826,00	1,63	191,82	0,34	162,40	0,19	130,26	0,16
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Deltaproteobacteria;D_3_Desulfovibrionales</i>	289,00	0,57	191,82	0,34	162,40	0,19	130,26	0,16
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Deltaproteobacteria;D_3_Myxococcales</i>	536,99	1,06	0,00	0,00	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria</i>	7 627,12	15,08	6 924,79	12,40	14 922,40	17,53	14 898,74	18,20
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Alteromonadales</i>	666,64	1,32	2 024,76	3,62	89,82	0,11	0,00	0,00
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Betaproteobacteriales</i>	166,08	0,33	56,69	0,10	142,75	0,17	165,18	0,20
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Cellvibrionales</i>	842,38	1,67	1 823,26	3,26	7 960,36	9,35	8 445,92	10,32
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_KI89A_clade</i>	112,17	0,22	0,00	0,00	114,91	0,14	106,57	0,13
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Nitrococcales</i>	0,00	0,00	0,00	0,00	263,70	0,31	257,78	0,31
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Nitrosococcales</i>	24,10	0,05	64,72	0,12	992,68	1,17	911,07	1,11
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Oceanospirillales</i>	5 381,87	10,64	2 665,92	4,77	2 601,18	3,06	2 551,92	3,12
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_OM182_clade</i>	347,53	0,69	172,01	0,31	2 522,55	2,96	2 362,09	2,89
<i>D_0_Bacteria;D_1_Proteobacteria;D_2_Gammaproteobacteria;D_3_Salinisphaerales</i>	86,35	0,17	0,00	0,00	0,00	0,00	0,00	0,00
<i>D_0_Bacteria;D_1_Verrucomicrobia</i>	147,32	0,29	67,21	0,12	240,44	0,28	199,14	0,24
<i>D_0_Bacteria;D_1_Verrucomicrobia;D_2_Verrucomicrobiae</i>	147,32	0,29	67,21	0,12	240,44	0,28	199,14	0,24
<i>D_0_Bacteria;D_1_Verrucomicrobia;D_2_Verrucomicrobiae;D_3_Verrucomicrobiales</i>	147,32	0,29	67,21	0,12	240,44	0,28	199,14	0,24

76 **Additional table 2: Abundance table of taxa found by the sequencing of the V3V4 16S region in the different samples of *Haslea ostrearia*.**

77 Null values have been grayed out. The mean depth was obtained using the ratio of the number of sequences aligned on each scaffold used for taxonomic annotation to the size
78 of these scaffolds; the percent abundance was calculated using the mean depth of the taxon relative to the total mean depth value (D_0_Bacteria)

79 **References**

- 80 1. P. J. Harrison, R. E. Waters, and F. J. R. Taylor, 'A broad spectrum artificial sea water medium
81 for coastal and open ocean phytoplankton', *Journal of Phycology*, vol. 16, no. 1, pp. 28–35, Mar.
82 1980, doi: [10.1111/j.0022-3646.1980.00028.x](https://doi.org/10.1111/j.0022-3646.1980.00028.x).
- 83 2. J. F. C. De Brouwer, K. Wolfstein, and L. J. Stal, 'Physical characterization and diel dynamics of
84 different fractions of extracellular polysaccharides in an axenic culture of a benthic diatom', *Euro.*
85 *J. Phycol.*, vol. 37, no. 1, pp. 37–44, Feb. 2002, doi: [10.1017/S0967026201003419](https://doi.org/10.1017/S0967026201003419).
- 86 3. C. Puppo, T. Voisin, and B. Gontero, 'Genomic DNA extraction from the pennate diatom
87 *Asterionella formosa* optimised for next generation sequencing. v1', preprint, Sep. 2017. doi:
88 [10.17504/protocols.io.jytcpxn](https://doi.org/10.17504/protocols.io.jytcpxn).
- 89 4. Z. Sun *et al.*, 'Community dynamics of prokaryotic and eukaryotic microbes in an estuary
90 reservoir', *Sci Rep*, vol. 4, no. 1, p. 6966, May 2015, doi: [10.1038/srep06966](https://doi.org/10.1038/srep06966).
- 91 5. P. M. Munyaka, N. Eissa, C. N. Bernstein, E. Khafipour, and J.-E. Ghia, 'Antepartum Antibiotic
92 Treatment Increases Offspring Susceptibility to Experimental Colitis: A Role of the Gut
93 Microbiota', *PLoS ONE*, vol. 10, no. 11, p. e0142536, Nov. 2015, doi:
94 [10.1371/journal.pone.0142536](https://doi.org/10.1371/journal.pone.0142536).
- 95 6. S. Andrews, 'FastQC: A Quality Control Tool for High Throughput Sequence Data'. 2010.
96 [Online]. Available : <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 97 7. P. Ewels, M. Magnusson, S. Lundin, and M. Källér, 'MultiQC: summarize analysis results for
98 multiple tools and samples in a single report', *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct.
99 2016, doi: [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354).
- 100 8. M. Martin, 'Cutadapt removes adapter sequences from high-throughput sequencing reads',
101 *EMBnet j.*, vol. 17, no. 1, p. 10, May 2011, doi: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).

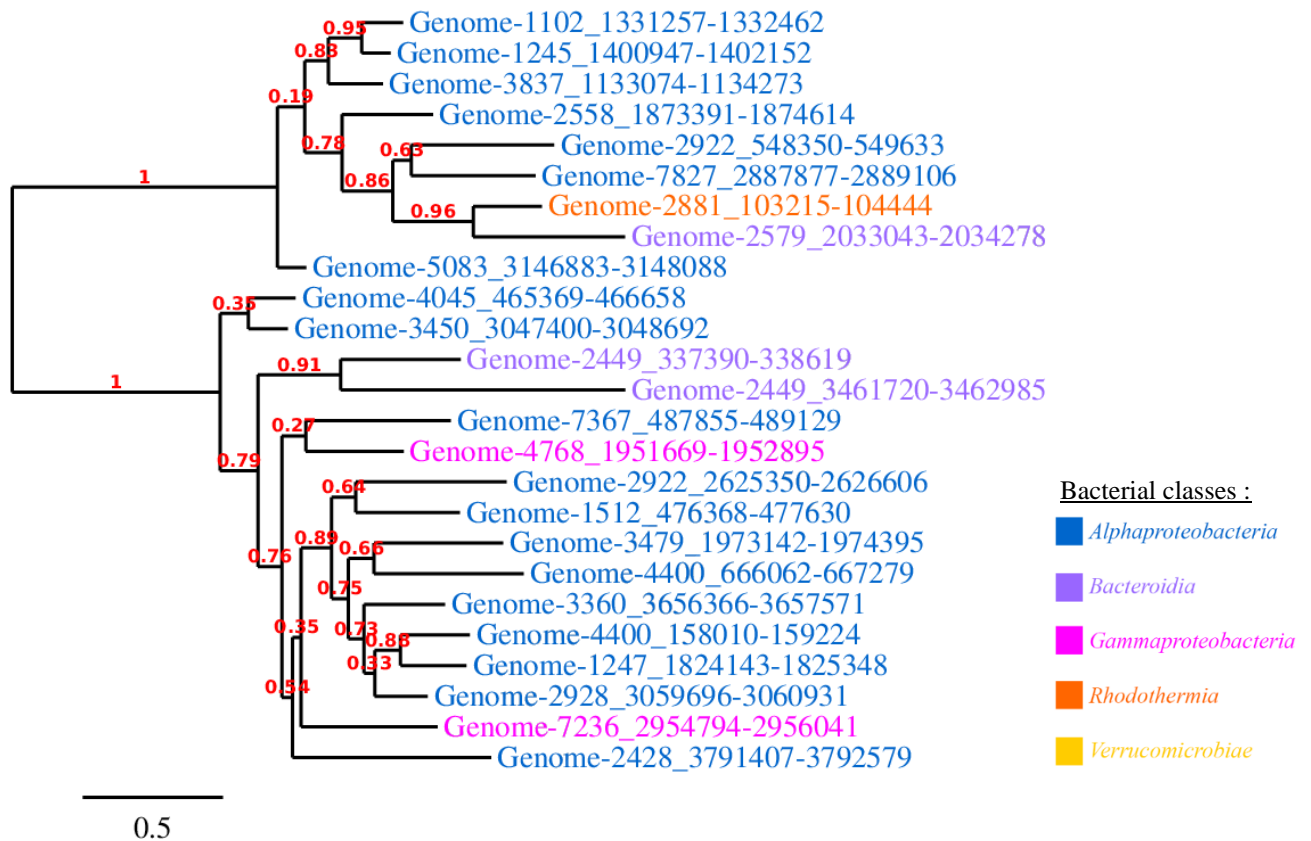
- 102 9. P. Pericard, Y. Dufresne, L. Couderc, S. Blanquart, and H. Touzet, 'MATAM: reconstruction of
103 phylogenetic marker genes from short sequencing reads in metagenomes', *Bioinformatics*, vol. 34,
104 no. 4, pp. 585–591, Feb. 2018, doi: [10.1093/bioinformatics/btx644](https://doi.org/10.1093/bioinformatics/btx644).
- 105 10. E. Bolyen *et al.*, 'Reproducible, interactive, scalable and extensible microbiome data science using
106 QIIME 2', *Nat Biotechnol*, vol. 37, no. 8, pp. 852–857, Aug. 2019, doi: [10.1038/s41587-019-0209-](https://doi.org/10.1038/s41587-019-0209-9)
107 [9](https://doi.org/10.1038/s41587-019-0209-9).
- 108 11. C. Quast *et al.*, 'The SILVA ribosomal RNA gene database project: improved data processing and
109 web-based tools', *Nucleic Acids Research*, vol. 41, no. D1, pp. D590–D596, Nov. 2012, doi:
110 [10.1093/nar/gks1219](https://doi.org/10.1093/nar/gks1219).
- 111 12. Md. Vasimuddin, S. Misra, H. Li, and S. Aluru, 'Efficient Architecture-Aware Acceleration of
112 BWA-MEM for Multicore Systems', in *2019 IEEE International Parallel and Distributed Processing*
113 *Symposium (IPDPS)*, Rio de Janeiro, Brazil, May 2019, pp. 314–324. doi: [10.1109/IPDPS.2019.00041](https://doi.org/10.1109/IPDPS.2019.00041).
- 114 13. P. Danecek *et al.*, 'Twelve years of SAMtools and BCFtools', *GigaScience*, vol. 10, no. 2, p.
115 giab008, Jan. 2021, doi: [10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008).
- 116 14. 'Python Language Reference'. Accessed: Dec. 15, 2022. [Online]. Available:
117 <http://www.python.org>
- 118 15. T. P. D. Team, 'pandas-dev/pandas: Pandas'. Zenodo, Nov. 22, 2022. doi:
119 [10.5281/ZENODO.3509134](https://doi.org/10.5281/ZENODO.3509134).
- 120 16. 'R: A language and environment for statistical computing'. Vienna, Austria, 2021. Accessed: Dec.
121 15, 2022. [Online]. Available: <https://www.R-project.org/>
- 122 17. H. Wickham, D. Navarro, and T. L. Pedersen, *ggplot2: Elegant Graphics for Data Analysis*.
123 Springer-Verlag New York, 2016. Accessed: Dec. 15, 2022. [Online]. Available:
124 <https://ggplot2.tidyverse.org>
- 125 18. H. Wickham, R. François, L. Henry, and K. Müller, 'dplyr: A Grammar of Data Manipulation'.
126 2022. Accessed: Dec. 15, 2022. [Online]. Available: <https://dplyr.tidyverse.org>

- 127 19. H. Wickham, 'Reshaping Data with the reshape Package', *J. Stat. Soft.*, vol. 21, no. 12, 2007, doi:
128 [10.18637/jss.v021.i12](https://doi.org/10.18637/jss.v021.i12).
- 129 20. H. Wickham, 'stringr: Simple, Consistent Wrappers for Common String Operations'. 2022.
130 Accessed: Dec. 15, 2022. [Online]. Available: <https://stringr.tidyverse.org>
- 131 21. E. Neuwirth, 'RColorBrewer: ColorBrewer Palettes'. Apr. 03, 2022. Accessed: Dec. 15, 2022.
132 [Online]. Available: <https://cran.r-project.org/package=RColorBrewer>

PREPRINT

133 4. Supplementary data of “Identification of antibiotic-resistance
 134 genes in order to reduce the bacterial content of *Haslea*
 135 *ostrearia* cultures”

136



137

138 **Additional Figure 1: Phylogenetic tree of the modified *tetA* gene from the bacterial genomes**
 139 **identified in *Haslea ostrearia* culture**

140 Each sequence is written in the format “Genome-*{ID}*_*{start}*_*{stop}*” which corresponds to the genome identifier
 141 in the metagenomic assembly, and the start and stop position of the gene of interest. The genes seen on the
 142 phylogenetic tree correspond to a modified *tetA*, which confers resistance to Tetracycline antibiotic. The leaves
 143 color indicates the class to which the bacterial genome from which each gene comes belongs. The phylogenetic
 144 tree was made by phylogeny.fr.

145

146

Genome ID in the assembly	Circular	Mean Depth	Size	GC content	Taxon_V1-V9	ARG detected by RGI vs CARD (80%ID, 70%cover)	"Resistance" Prokka genes, all(unique)	Busco score vs class DB	BioSample
1102	N	100	3185041	57.35	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Marivita	2	13(11)	80.6	SAMN33565282
1245	N	114	3183699	57.61	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Marivita	1	14(12)	65.3	SAMN33565283
1247	N	160	3038566	57.88	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Marivita	1	15(13)	58.6	SAMN33565284
1512	Y	72	3500668	62.25	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Sphingomonadales;D_4__Sphingomonadaceae;D_5__Erythrobacter	3	14(9)	93.8	SAMN33565285
1514	Y	1969	3093397	53.59	D_0__Bacteria;D_1__Verrucomicrobia;D_2__Verrucomicrobiae;D_3__Verrucomicrobiales;D_4__Rubritaleaceae;D_5__Luteolibacter;D_6__uncultured_bacterium	2	8(6)	96.4	SAMN33565286
1677	Y	248	3621925	56.54	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Cellvibrionales;D_4__BD2-7	7	20(13)	97.3	SAMN33565287
2428	Y	324	4507981	59.52	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhizobiales;D_4__Rhizobiaceae;D_5__Hoefflea	7	24(14)	99.5	SAMN33565288
2449	Y	2961	4167685	41.67	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Crocinitomicaceae;D_5__Fluviicola;D_6__uncultured_bacterium	2	8(6)	63.4	SAMN33565289
2558	Y	156	5005768	64.74	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Mameliella	6	19(12)	86.1	SAMN33565290
2579	Y	126	3921997	38.04	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Flavobacteriaceae;D_5__Maribacter	0	16(10)	82.7	SAMN33565291
2648	Y	79	3969919	60.23	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae	3	17(12)	90.2	SAMN33565292
2661	Y	356	3506103	58.66	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Oceanospirillales;D_4__Alcanivoracaceae;D_5__Alcanivorax	5	27(18)	92.6	SAMN33565293
2753	Y	146	4225817	33.90	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Flavobacteriaceae	3	9(9)	84.0	SAMN33565294
2879	N	332	4374826	60.88	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Cellvibrionales	4	26(13)	93.9	SAMN33565295
2881	N	29	2225183	35.53	D_0__Bacteria;D_1__Bacteroidetes;D_2__Rhodothermia;D_3__Balneolales;D_4__Balneolaceae;D_5__Balneola	3	10(7)	54.9	SAMN33565296
2922	Y	397	3400785	56.70	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Caulobacterales;D_4__Hyphomonadaceae;D_5__uncultured	3	13(10)	94.5	SAMN33565297
2928	Y	155	3497831	61.19	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Roseovarius	4	13(10)	97.7	SAMN33565298
3360	Y	142	3790680	62.26	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae;D_5__Sulfitobacter	1	14(10)	78.9	SAMN33565299
3450	Y	419	3911877	59.38	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhizobiales;D_4__Rhizobiaceae;D_5__Cohaesibacter	4	21(14)	96.8	SAMN33565300

3479	Y	3104	2970267	61.92	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Caulobacterales;D_4__Hyphomonadaceae; D_5__Oceanicaulis	1	13(10)	98.1	SAMN33565301
3502	Y	137	4872944	39.81	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Cytophagales;D_4__Cyclobacteriaceae; D_5__Marinoscillum	1	14(5)	85.5	SAMN33565302
3837	Y	404	3984771	60.14	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae	3	11(9)	99.3	SAMN33565303
4045	Y	188	3450511	57.26	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhizobiales;D_4__Rhizobiaceae	2	12(10)	99.3	SAMN33565304
4400	Y	210	3658206	57.92	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae; D_5__Loktanella	3	17(11)	99.5	SAMN33565305
4768	Y	967	4126287	53.99	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__OM182_clade	5	27(16)	94.5	SAMN33565306
5083	Y	815	4378345	57.26	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae	4	16(11)	87.1	SAMN33565307
6916	N	25	2871272	39.82	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Sphingobacteriales;D_4__NS11-12_marine_group	2	11(8)	72.5	SAMN33565308
7192	N	55	3305667	52.31	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Burkholderiaceae; D_5__Limnobacter	6	21(11)	92.1	SAMN33565309
7236	Y	42	4622411	52.98	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__KI89A_clade	5	35(20)	92.8	SAMN33565310
7367	Y	420	3444699	58.84	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria	2	22(16)	99.0	SAMN33565311
7827	Y	104	2955266	59.57	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Caulobacterales;D_4__Hyphomonadaceae; D_5__Hyphobacterium	4	16(12)	96.7	SAMN33565312

147 **Additional Table 1: Summary table of the bacterial genomes identified in *Haslea ostrearia* culture**

148 The information given comes from the statistics created by the Flye software during the metagenomic assembly of the genomes. The taxonomic report was obtained using QIIME2 and

149 SILVA database 138 Ref NR 99.

150

151

152

153

154

Genome ID	Aligning the 1000 bases before the start of the gene	Aligning the 1000 bases after the end of the gene	Gene 1 (positions and strand)	Gene 2 (positions and strand)
1677	No similarity found	98%ID, 47 aligned bases	488668-489891(+)	505781-507004(+)
2428	No similarity found	No similarity found	4348606-4349781(-)	4368038-4369213(-)
2648	No similarity found	No similarity found	1661439- 1662614(+)	3890687-3891862(-)
2661	No similarity found	No similarity found	2896221- 2897411(+)	2912985-2914175(+)
2879	No similarity found	95%ID, 44 aligned bases	779478-780701(+)	799994-801217(+)
2922	100%ID, 58 aligned bases	No similarity found	71719-72909(-)	2892216-2893406(+)
2928	No similarity found	100%ID, 45 aligned bases	998623-999612(-)	1041109-1042284(-)
3450	No similarity found	No similarity found	761208-762383(+)	849198-850373(+)
3837	No similarity found	No similarity found	1316494-1317669(-)	1336956-1338131(-)
4045	No similarity found	No similarity found	1868115- 1869290(+)	1887439-1888614(+)
4768	No similarity found	No similarity found	300623-301816(-)	317194-318243(-)
5083	No similarity found	100%Id, 44 aligned bases	269649-270824(+)	3216203-3217378(-)
7192	No similarity found	No similarity found	308005-309198(-)	324351-325544(-)
7236	No similarity found	No similarity found	2395054-2396238(-)	2411667-2412851(-)
7367	No similarity found	No similarity found	861857-863047(+)	879076-880266(+)
7827	No similarity found	98%ID, 80 aligned bases	1033348- 1034523(+)	1053234-1054409(+)

156 **Additional Table 2: Results of alignments between the extremities of *EF-Tu* genes found in**
157 **bacterial genomes**

158 The extremities for which several copies of the *EF-Tu* gene were found in the same genome were aligned by Blastn
159 in default mode with each other according to their position relative to the gene. The 1000 bases before the start of
160 the gene and the 1000 bases after the end of the gene were compared in pairs, the result is given according to the
161 information obtained by Blast (“No similarity found” or the percentage of identity as well as the number of aligned
162 bases). The genes found in each genome were annotated as follows in the table: start position of the gene - end
163 position of the gene (strand).

164

Titre : Caractérisation in silico des bactéries des cultures de la microalgue marine *Haslea ostrearia*

Mots clés : *Haslea ostrearia*, Mitochondrie, Chloroplaste, Bactéries, Identification taxonomique, Antibiorésistance, Génomique, Métagénomique

Résumé : *Haslea ostrearia* est une algue microscopique capable de synthétiser un pigment bleuté, la marennine, très prisé dans de nombreux domaines suite à ses applications potentielles multiples (biocarburant, colorant alimentaire, compléments en aquaculture, etc.). Cependant, bien que connue depuis 1820, les informations disponibles à ce jour sur *Haslea ostrearia* (Gaillon) Simonsen restent faibles. Suite à la complexité de cultiver cette dernière dans des conditions axéniques, aucun génome n'est disponible pour cette microalgue dans les bases de données actuelles. C'est pourquoi, dans un premier temps, cette thèse a extrait le génome de ses organelles de l'assemblage des séquençages métagénomiques obtenus des cultures de *Haslea ostrearia*, tout en mettant en

avant l'importante concentration bactérienne des données ADN empêchant d'obtenir le génome nucléaire. Les travaux se sont ensuite concentrés sur l'identification des bactéries présentes dans les cultures de *Haslea ostrearia* et sur la nécessité de traiter ces dernières pour assurer leur reproductibilité. Les génomes bactériens ont également été étudiés afin de dresser la liste des antibiotiques capables de réduire la diversité bactérienne dans les cultures en recherchant les gènes codant pour des résistances aux antibiotiques et aux antimicrobiens. Cette thèse a permis de mieux comprendre les organismes peuplant les cultures de *Haslea ostrearia*, mais également de proposer des pistes pour l'amélioration du protocole de culture actuellement utilisé, afin d'obtenir le génome de la microalgue.

Title : In silico characterisation of bacteria from cultures of the marine microalga *Haslea ostrearia*

Keywords : *Haslea ostrearia*, Mitochondria, Chloroplast, Bacteria, Taxonomic identification, Antibiotic resistance, Genomics, Metagenomics

Abstract : *Haslea ostrearia* is a microscopic alga capable of synthesising a bluish pigment, marennin, which is highly valued in many fields because of its many potential applications (biofuel, food colouring, aquaculture supplements, etc.). Although known since 1820, information on *Haslea ostrearia* (Gaillon) Simonsen is still scarce. No genome for this microalga is available in current databases due to the complexity of growing it under axenic conditions. Therefore, this work first extracts the genome of its organelles from the metagenomic sequencing assembly obtained from *Haslea ostrearia* cultures, highlighting the high bacterial concentration of DNA data that prevents

obtaining the nuclear genome. Work then focused on the identification of the bacteria present in *Haslea ostrearia* cultures and the need for treatment to ensure their reproducibility. Bacterial genomes were also studied to identify antibiotics that could reduce bacterial diversity in the cultures by searching for genes that code for resistance to antibiotics and antimicrobials. This work has enabled us to gain a better understanding of the organisms that populate *Haslea ostrearia* cultures, but also to suggest ways of improving the culture protocol currently used to obtain the genome of the microalga.